Engagements with value in cancer clinical research design: Prioritizing and planning beyond the public

Andrew Staver Hoffman
Department of Sociology
McGill University, Montréal
April 2019

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Sociology.

A fantasy (at any rate, what I call a fantasy): a resurgence of certain desires, certain images that lurk within you, that you want to be identified by you, sometimes your whole life, and often only assume concrete form thanks to a particular word. That word, a key signifier, is what leads from the fantasy to its investigation. To mine the fantasy through snatches of knowledge=research. [...]

Now, it was in the course of a chance reading [...] that the fantasy encountered the word that would set it to work.

Roland Barthes,
How to Live Together

Acknowledgments

As it turns out, a big problem with taking a while to write a dissertation isn't just its possible sequelae for one's career prospects. It's also the fact that one accrues an ever-growing number of individuals to whom they are indebted, rendering the list not just excessively long, but also inevitably incomplete -- the result of a brain that grows progressively more forgetful with each passing academic year, hampering recall all of those who have helped along the way. Needless to say, countless individuals and organizations have contributed to the successful completion of this thesis, extending (oftentimes combinations of) intellectual, financial, institutional, material, and psychosocial support to me over the course of my research and writing process. I will do my best to acknowledge them here, and apologies to all whom I have forgotten in the process of trying to remember.

It goes without saying that the unflinching mentorship and camaraderie of my doctoral supervisor, Alberto Cambrosio, has been central here. This began in my very earliest days as a doctoral student in McGill's sociology program, where his courses in social studies of medicine furnished me with a solid foundation in the sociology of biomedical knowledge, and during which time he also encouraged me to explore clinical practice guidelines as an important site for unpacking the production and evaluation of biomedical knowledge. Alberto's guidance continued, unabated, even as the empirical focus of my doctoral research shifted -- from guidelines to comparative effectiveness research -- as did my analytic commitments as I worked to sort out a series of concerns pertaining to stakeholder engagement, economic valuation, and the question of the 'public.' These changes also precipitated what would become my extensive engagement with French *sociologie pragmatique*, a body of literature I am grateful to Alberto for first introducing me to very early on in my postgraduate career, and for which he continues to serve as provocative interlocutor, constantly pushing me to grow.

It was not until several years into my studies that I first encountered Laura Centemeri, whose pragmatic sociology-inspired work I had read and whom I contacted on a whim at the suggestion of a mutual colleague; cheers, Martin Giraudeau, for planting that seed. This quickly evolved into an exciting email exchange, followed thereafter by a nearly weekly Skype chat ritual that lasted the better part of a year (and still continues today, although regrettably less frequently.) Laura helped me to understand the intellectual roots and historical context in which Laurent Thévenot's *régimes d'engagement* framework emerged, enabling me to see ever more clearly its utility for my own research. She was a dedicated listener as I talked out most of the germinal ideas of the following chapters. Our

conversations sparked a period of intense evolution in my style of thinking about the world; they proved indispensable for conceptualizing the peculiar modes of coordination I had observed in my fieldwork, but which I had struggled for some time to place in conversation with the existing sociological scholarship on participation and quantification. More to the point, I may have never gotten to write these acknowledgments without the informal tutelage she provided me over the course of writing up; the French translation of the thesis abstract also benefited from her insights.

My gratitude also goes to the many individuals working in and around the comparative effectiveness research milieu who entertained my sometimes rambling interview questions; were kind enough to let me observe and participate in their meetings and workshops; and shared with me documents that were not always publicly available. Particular thanks is due to the CANCERGEN project organizers and External Stakeholder Advisory Group members with whom I had the pleasure of interacting on numerous occasions over the course of my fieldwork.

This project was the beneficiary of a number of different funding sources. This included doctoral training fellowships with the Healthcare, Technology, and Place program at the University of Toronto and the Quebec-based APOGÉE-Net/CanGèneTest Network of Investigators and Decision Makers. My exposure to the group of interdisciplinary colleagues which accompanied my involvement in these programs -- in particular, Peter Coyte, Eric Mykhalovskiy, and Anne Louise Pontigon Montgomery at HCTP and Renaldo Battista at APOGÉE-Net/CanGèneTest -- were instrumental for the development and diversification of my knowledge about evaluative research in healthcare and the financial support these programs provided allowed me to focus on my research full-time.

I was also awarded graduate student research travel awards by McGill University and the Social Science and Humanities Research Council of Canada, enabling me to attend meetings and conduct interviews as part of my ongoing fieldwork. A funded stay at the Brocher Foundation in Hermance, Switzerland provided two months of dedicated writing time which I had the good fortune of passing in beautiful lakeside surroundings, and in the company of a keen group of international colleagues; thinking through the scope of my own project greatly benefitted from formal and informal engagements with their work, which addressed all manner of social, ethical, and legal elements of health policy and biomedical technology. McGill University's Wolfe Graduate Fellowship in Scientific and Technological Literacy provided an extra bit of funding during this writing up phase.

At McGill, my professors in the sociology department shaped me as a budding sociologist, while my many friends made sure that graduate school was never an isolating experience. The list would be too long to recount in full, but deserving special mention are my departmental pals and sometime-roommates Zygmunt Mochnacki and Nicole Denier; and my SSOM-ish buddies Clarissa

Allen, Sarah Berry, Jonah Campbell, Emilio Dirlikov, Jessica Dolan, Hadi Karsoho, Hanna Kienzler, Loes Knaapen, Pierre Minn, Raul Necochea, Nicole Nelson, Eugene Raikhel, Susan Rogers, and Martina Schlünder.

Beyond McGill, my studies also benefited from visiting studentships at the Institute for Advanced Study on Science, Technology and Society in Graz, Austria; the department of Anthropology at New York University; and the ValueS group in the department of Thematic Studies at Linköping University in Sweden. Regular participation in these international scholarly communities was always a pleasure and initiated many intellectual bonds whose influence on my academic trajectory continue to resonate today: Baki Cakici, Allison Cool, Véra Ehrenstein, Lotta Björklund Larsen, Francis Lee, Dick Magnussen, Emily Martin, Laura Mauldin, Amber Nelson, Rayna Rapp, and Emily Yates-Doerr are just a few of the individuals who helped facilitate or otherwise contributed to these stays.

CF Helgesson fundamentally changed my career by opening my eyes to the emerging field of valuation studies, while his work on the valuations informing the design of clinical trials has provided much common ground for our many chats over the years. Stuart Hogarth has been a wonderful source of knowledge about biomedical regulation, and the regulation of diagnostic technologies in particular, in a field where this has been a mostly neglected topic; he was also instrumental in securing me access to key documents pertaining to my research, without which large parts of this analysis would not have been feasible. Tiago Moreira's work on clinical practice guidelines was important for my own early interest in this area, while his subsequent work on evaluative technologies in health economics became a key theme of our recurring conversations in the ensuing years as my own work on CER evolved. Teun Zuiderent-Jerak was another early academic role model, practically since week one of my doctoral studies when Alberto tossed me his PhD thesis and instructed me to read it posthaste; he would eventually morph into a dear friend -- never at a loss for a kind or encouraging word to share -- and I continue to look up to him, both personally and professionally.

Thanks, also, to the many individuals with whom I have shared conference and workshop panels. In particular, I would like to acknowledge Stefan Timmermans' incisive comments on an early version of Chapter 3, which he provided at Columbia University's 'Expertise from Margin to Center' conference in 2016, as well as Laurent Thévenot's own helpful commentary on a subsequent draft of the same chapter, also delivered in 2016 at the 'Politics of Engagement in an Age of Differing Voices' workshop in Helsinki.

My current research team -- mentors David Ribes and Geof Bowker and fieldwork coequal Steve Slota -- have demonstrated great patience and flexibility while providing me with new possibilities for expanding my empirical and theoretical knowledge base. My Data Ecologies Lab-mates in the

department of Human Centered Design and Engineering at the University of Washington have similarly allowed me to see social studies of infrastructure from fresh perspectives. I appreciate the many opportunities and chances you have all lent me as I further my growth as a student of socio-technical systems.

My family, both blood and chosen, have been an enduring source of support throughout my life and career. In New York, my mother, Robin; father, Michael; and my Aunt Root and Uncle Mike always offer at least a hot meal or a couch to crash on, if not my very own bedroom for extended stays. The same is true for my sister Dori Jill, brother-in-law Elli, and niece Harlow Sloane in Connecticut. In Montréal, Kelly McKinney and Bruce and Bea Mellett have always welcomed me with open arms, as have Gerard, Mimi, and Fredou Le Nezet in France. Each of these individuals, along with my many friends near and far, have molded me in ways both subtle and obvious; for this I am eternally grateful. My two grandmothers, Bertha Hoffman and Sally Staver, always pushed me to excel in whatever it was that I did and would have loved to hold in their hands a copy of my completed thesis. My procrastination unfortunately outlasted their biological endurance and so it is to both of their memories that I dedicate this whole endeavor.

But there is probably no living person more instrumental to my thriving -- and more relieved to see this project through to its conclusion -- than my dear partner, Elo Le Nezet. Any attempt to conjure words that would adequately express my gratitude for you would be an exercise in futility. Sure, you have translated the abstract for this thesis into French, and this is of course a very practical contribution. But at a more fundamental level, for the past four years you've nourished me with care and, in the roughest times, have literally fed and housed me. You've sacrificed holidays and moved across the country for me. You've said 'Yes' to two marriage proposals but have smartly resisted tying the knot until I can rightfully add a couple of letters before my name and a few more at the end. And now, at long last, with the good graces of my thesis committee, I look forward to the many years of nuptial bliss that our dissertation-free future promises. Thank you for being you. I love you.

Table of contents

List of figures & tables

Abstract

Comparative effectiveness research (CER) first became a matter of national policy concern in the United States around the year 2009 as a novel approach to evaluating the utility of healthcare interventions. Spurred on by over US$1 billion of federally apportioned CER funding, the development of infrastructures and methods for conducting comparative effectiveness research swiftly followed, as did their uptake in nearly every corner of biomedicine. There was, however, particular fervor within the oncology community, the medical subfield specializing in cancer care and prevention: this was a prime site of innovation in genomic diagnostics and targeted therapy, but also a domain in which practitioners and policymakers were openly questioning the clinical usefulness and cost-effectiveness of these technologies. Comparative effectiveness research thus surfaced as an attractive means for arbitrating how the rapidly appearing tools of personalized medicine measured up against existing standards of care.

The initial point of departure for this thesis is the emergence and instantiation of CER in the US. Deploying ethnographic and documentary research methods, and focusing primarily on the case of the Center for Comparative Effectiveness Research on Cancer Genomics (CANCERGEN), it traces the myriad coordination and (e)valuation practices which collectively came to constitute the work of organizing, prioritizing, and designing CER studies of personalized medicine technologies in oncology. Importantly, the analysis progressively reveals a set of characteristics that were shared across the many elements of CANCERGEN's unique enactment of CER -- namely, in how project organizers devised its innovative stakeholder engagement and economic decision modeling techniques -- which, taken together, point away from what might be qualified as *public* coordination mechanisms and instead signify a more limited *planning* orientation.

Turning to recent work at the nexus of French *sociologie pragmatique* and science & technology studies, the thesis goes on to outline an ecological, temporally relational approach to (e)valuation. By considering this *ecology of engagements and (e)valuations*, it enhances the analyst's capacities for recognizing prima facie the existence of *pre-public* modes of coordination and their corresponding evaluative formats. Second, it is propitious to a more faithful accounting of the features that facilitate or otherwise preclude situations and artefacts from transiting between moments of public legibility and those of a more confined nature. These contributions address a significant lacuna in current sociological scholarship on evaluation, participation, and quantification. Namely, they push further towards the development of a comprehensive analytic lexicon that is vitally needed for unpacking the drivers and consequences of both public and other-than-public ways of engaging with value.

Résumé

La «comparative effectiveness research» (CER) est tout d'abord devenue une question de préoccupation de politique nationale aux Etats-Unis aux alentours de l'année 2009 comme une nouvelle approche pour évaluer l'utilité des interventions du système de santé. Sous l'impulsion de plus de $1 milliard de dollars US de fonds fédéraux répartis pour la CER, un développement des infrastructures et méthodes de conduite de comparative effectiveness research s'en est rapidement suivi, ainsi que leur absorption dans presque chaque recoin de la biomédecine. Il y a eu, cependant, une ferveur particulière au sein de la communauté de l'oncologie, le sous-domaine médical spécialisé en soin et prévention du cancer: ce dernier était un site majeur d'innovation en matière de diagnostics génomiques et de thérapie ciblée, mais aussi un domaine dans lequel les praticiens et les décideurs questionnaient ouvertement l'utilité clinique et la rentabilité de ces technologies. La "comparative effectiveness research" a ainsi fait surface comme un moyen attractif d'arbitrer la façon dont ces outils de médecine personnalisée rapidement apparus se mesuraient aux standards de soin déjà existants.

Le point de départ initial de cette thèse est l'émergence et l'instanciation de la CER aux Etats-Unis. Déployant des méthodes de recherche ethnographique et documentaire, et se concentrant principalement sur le cas du «Center for Comparative Effectiveness Research in Cancer Genomics» (CANCERGEN), elle trace une myriade de coordination et de pratiques d'(é)valuation qui sont collectivement venues à constituer la tâche d'organiser, de donner la priorité et de concevoir les études de la CER sur les technologies de médecine personnalisée en oncologie. De manière importante, l'analyse révèle progressivement un ensemble de caractéristiques qui étaient partagées parmi les nombreux éléments de promulgation unique de CER de CANCERGEN - à savoir la façon dont les organisateurs de projet ont conçu la participation innovatrice des intervenants ainsi que les techniques de modélisation de décision économique - ce qui, pris ensemble, pointe loin de ce qui pourrait être qualifié de mécanismes de coordination *publique* et qui au contraire signifie une orientation de *planification* plus limitée.

Se tournant vers des travaux récents à la croisée de la *sociologie pragmatique* française et des études en science et technologie, la thèse continue à esquisser une approche écologique, temporellement relationnelle à l'(é)valuation. En considérant cette *écologie des engagements,* cela accroît les capacités de l'analyste à reconnaître à première vue l'existence de modes *pré-publics* de coordination et leurs formats d'évaluation correspondants. En outre, il est propice à rendre compte de manière plus fidèle des caractéristiques qui facilitent ou, par ailleurs, empêchent des situations ou artefacts de transiter

entre des moments de lisibilité publique et d'autres de nature plus confinée. Ces contributions abordent une lacune considérable dans les connaissances sociologiques actuelles sur l'évaluation, la participation et la quantification. C'est-à-dire, elles poussent plus loin vers le développement d'un lexique analytique complet qui est absolument nécessaire afin de mieux saisir les incitations ainsi que les conséquences des manières à la fois publiques et autres-que-publiques de tester, attester, et contester la valeur dans des dispositifs d'(é)valuation impliquant la participation.

Preface

This manuscript-based thesis represents the culmination of a singular research program focused on the emergence and enactment of comparative effectiveness research towards which the study design, collection and analysis of data were oriented. Its four empirical chapters (Chapters 2-5) have been written such that they are suitable as standalone scholarly publications, e.g. peer-reviewed journal articles or chapters of an edited volume. The reader may thus notice slight overlap and repetition of empirical contextualization between several of the chapters, an artefact of the thesis format itself.

I am the sole author of all thesis contents save for Chapter 2, 'Comparative Effectiveness Research in Health Technology Assessment,' where I acted as primary author. My two co-authors, Alberto Cambrosio and Renaldo Battista, provided invaluable insights and feedback on its contents and structure. The chapter has been published in the Springer *Health Services Research* series special volume on *Comparative Effectiveness Research in Health Services*, edited by Adrian Levy and Boris Sobolev. The remaining chapters are currently under preparation for submission to peer-reviewed journals.

To the best of my knowledge, this thesis stands as the only existing qualitative sociological study of comparative effectiveness research writ large, and the sole ethnographic account of the Center for Comparative Effectiveness Research in Cancer Genomics. As such, it makes an important empirical contribution to the wider corpus of scholarship within and beyond medical sociology. Moreover, the interventions made throughout the chapters that follow provide a number of original conceptual insights to the (sub)fields of economic sociology, science and technology studies, the sociology of biomedicine, and valuation studies, pushing further towards the development of a more comprehensive analytic lexicon which cuts across sociological scholarship on participation, quantification, and wider theorizing on the public and its other(s).

Glossary

ACS: American Cancer Society (US)

AHRQ: Agency for Healthcare Research and Quality (US)

AMA: American Medical Association (US)

ARRA: American Recovery and Reinvestment Act

BC: Breast Cancer

BIA: Budget Impact Analysis

BRAF: Proto-oncogene B-Raf (gene)

BRCA: Breast Cancer Susceptibility Gene

BTT: biomarker/treatment trials

CANCERGEN: Center for Comparative Effectiveness Research in Cancer Genomics

CBA: Cost-Benefit Analysis

CCA: Cost-Consequence Analysis

CEA: Cost-Effectiveness Analysis

CED: Coverage with Evidence Development

CER: Comparative Effectiveness Research

CGP: Cooperative Group Program

CLIA: Clinical Laboratory Improvement Amendments

CMPT: Center for Medical Technology Policy

CMS: Centers for Medicare and Medicaid Services (US)

CPG: Clinical Practice Guidelines

CT: Computed Tomography

CTA: Constructive Technology Assessment

CTEP: Cancer Therapy Evaluation Program

CUA: Cost-Utility Analysis

DFS: Disease-Free Survival

DNA: Deoxyribonucleic Acid

EGFR: Epidermal Growth Factor Receptor

ELSI: Ethical, Legal, and Social Implications

EMR: Electronic Medical Record

ERCC1: ERCC Excision Repair 1 (gene)

ESAG: External Stakeholder Advisory Group

EUnetHTA: European Network for Health Technology Assessment

EVPI: Expected Value of Perfect Information

EVPPI: Expected Value of Partially Perfect information

EVSI: Expected Value of Sample Information

FCC-CER: Federal Coordinating Council for Comparative Effectiveness Research (US)

FDA: Food & Drug Administration (US)

G-BA: Federal Joint Committee (Germany)

GEP: Gene Expression Profiling

GO Grant: Grand Opportunities Grant

GPM: Genomics and Personalized Medicine

HAS: National Authority for Health (France)

HHS: Department of Health & Human Services (US)

HR: Hormone Receptor

HTA: Health Technology Assessment

HTAi: Health Technology Assessment International

ICER: Incremental Cost Effectiveness Ratio

INAHTA: International Network of Agencies for Health Technology Assessment

IOM: Institute of Medicine

KRAS: Oncogene in Kirsten RAt Sarcoma Virus

LDT: Laboratory Developed Test

MA: Marginal Analysis

MAS: Medical Advisory Secretariat (Ontario)

MMA: Medicare Modernization Act

NCI: National Cancer Institute (US)

NCTN: National Clinical Trials Network

NHS: National Health Service (UK)

NICE: National Institute for Health and Care Excellence (UK)

NIH: National Institutes of Health (US)

NSCLC: Non Small Cell Lung Cancer

OHTAC: Ontario Health Technology Assessment Committee (Canada)

OTA: Office of Technology Assessment (US)

PCOR: Patient Centered Outcomes Research

PCORI: Patient Centered Outcomes Research Institute (US)

PI: Principal Investigator

PM: Personalized Medicine

PPACA: Patient Protection and Affordable Care Act

PPI: Patient and Public Involvement

PRO: Patient Reported Outcome

QALY: Quality-Adjusted Life Year

RCT: Randomized Controlled Trial

RNA: Ribonucleic Acid

ROE: Regimes of Engagement

STS: Science & Technology Studies

SWOG: Southwest Oncology Group

TTP: Test Target Profile

VOI: Value of Information

VOR: Value of Research

VPI: Value of Perfect Information

VSI: Value of Study Information

WoS: Web of Science

WTP: Willingness-to-Pay

Chapter 1

Engagements with value in comparative effectiveness research: an introduction

1. On encountering a new field of inquiry in medical research

Who and what gets to count in the world? And how can we best use the conceptual tools of sociological inquiry to understand the practical as well as political implications bound up in asking and answering this question?

Such concerns motivate the arguments contained in the present thesis: an empirical study of the emergence of an ostensibly novel type of health research - referred to as comparative effectiveness research - and an organization called the Center for Comparative Effectiveness Research in Cancer Genomics (henceforth CANCERGEN), which I situate as an exemplar of this particular research approach. At the heart of my analysis are the manifold processes whereby things (human and nonhuman alike) are made to count - some less, others more - in defining, prioritizing, and designing clinical research studies that evaluate the utility of diagnostic technologies in cancer medicine. To speak of people and objects being 'made to count,' I intentionally gesture to the double meaning of this phrase: that is, in the sense of someone or something receiving the command or authority to *act as arbiter of the valuable* while at the same time remaining vulnerable to being a *subject of* that very evaluative work.

In analyzing early debates around how the concept and contents of comparative effectiveness research should be defined, along with close ethnographic analyses of CANCERGEN and its technologies of stakeholder participation and economic modeling, we find a curious admixture of evaluative constructs. However, what we also find amidst this landscape are coordinative assemblages that, perhaps to the surprise of the analyst, resist their escalation into a realm of *publicity*. This thesis thus contributes simultaneously to multiple strands of theoretical and empirical reflection in the fields of sociology and science and technology studies, including (1) the politics of participation in science and medicine; (2) the processes and practices of quantification and economization; and (3) wider discussions around what it means to for something to qualify as a

public phenomenon. In turn, it proposes a novel means for addressing the affordances and limitations of what I label *pre-public* phenomena. By this I intend to signal a certain situational parsimony wherein actors may reflexively acknowledge the existence of a public world beyond their more immediate settings; nevertheless, they resist extrapolating their activities into these extant spaces and future moments, instead orienting coordination around more proximal concerns and towards local demonstrations of utility.

~

Before unpacking what is meant by these terms, though, it begs noting that my empirical interest in the topic of comparative effectiveness research as well as my conceptual interest in disentangling public(s) and their other(s) (e.g. Berg & Timmermans 2000) are outcomes of my research process rather than its starting point. My preliminary doctoral fieldwork was in fact focused on the landscape of organizations in the United States that create and publish clinical practice guidelines (CPGs) in oncology (cf. Knaapen et al. 2010; Moreira 2005; Weisz et al. 2007). These are summary documents that provide physicians and other clinical staff with recommendations for how to treat a range of medical conditions; guidance is typically based on results from previously conducted clinical studies, with formalized syntheses and appraisals of randomized controlled trials (RCTs) -- referred to as *systematic reviews* or *meta-analyses* -- standing as the 'gold standard' of biomedical evidence (Bohlin 2012; Timmermans & Berg 2003). My knowledge of this domain predated my move to Montréal, where I eventually relocated to begin my doctoral degree. During the couple of years prior to applying to PhD programs, I worked as a research data manager in the clinical trials office at a large academic cancer center in New York City. Knowing that guideline development often entailed convening interdisciplinary panels of clinicians and clinical research experts (and sometimes patients, too), I was keen to learn more how these groups go about grading the quality of evidence in a given disease area and arbitrating what they considered to be legitimate knowledge in deriving clinical guidance documents.

As it so happened, I found myself living back in New York City during the summer of 2009 and had taken up a part time evening position filling out case report forms for industry-sponsored cancer clinical trials that were being conducted at the hospital where I had previously worked. The hospital in question was also part of a large national Consortium, comprised of the top 21 academic cancer centers in the US. Faculty and staff from the Consortium member institutions volunteer their

time serving on a number interdisciplinary disease-specific panels, made up of experts from the three modalities of cancer treatment -- medical oncology, radiation oncology, and surgical oncology. What results from these standing committees is a suite of treatment guidelines that cover most forms of cancer. These particular guidelines, which are translated into a number of languages, are some of the most well-known and widely used guidelines, both in the US and internationally,

The thought occurred to me that I was sitting on a treasure trove of ethnographic data. All I needed to do was find out who at my hospital was active on these panels and send some emails around politely requesting interviews. The email I sent to Dr. Forrester (a pseudonym), a highly regarded oncologist at the hospital, would be my first go at this data collection strategy. I had come across his name on the Consortium's website, where I learned that he was not only a member of several different guideline panels but was also a member of the Consortium-wide guideline steering committee. So off went my email requesting an interview, which I sent from my official hospital email address -- expecting that this would somehow lend me an air of authority, or at the very least reduce any suspicions he might have about a junior sociologist showing up at his office to ask a list of detailed questions about the knowledge synthesis and evaluation practices in which he participated. I was pleased to find in my inbox an affirmative reply to my request shortly thereafter. And so it was around 3pm on a warm and humid Friday afternoon mid-August that, armed with a digital recorder, a notepad, and a blank consent form, I excitedly arrived to Dr. Forrester's Upper East Side office for my scheduled meeting.

I had done my best to explain the scope of my research in my initial email to Dr. Forrester, which I repeated again at the start of our meeting. It quickly occurred to me, however, that the situation wasn't quite what he was expecting. A methodological memo that I wrote up immediately following the interview recounts that

> [Dr. Forrester seemed] caught off guard by the formalities of the interview. First, he assumed based on the 'flavor' of my email that I would not require much time to meet with him, so his secretary had only scheduled enough time for a 'chat.' As well, upon my handing him a consent form to sign, he expressed some hesitation and said that I would have to return the following Friday so that he would have some time to read [it] over [...] When I returned for the actual formal interview, he handed me the consent form. He asked if he had the right not to be tape recorded, and as I was replying that, [yes], he

> did have that right - [although recording] makes it easier for me as a researcher - he nodded his head somewhat dismissively as he handed back the signed consent.

I recall feeling quite deflated by this experience. At times, Dr. Forrester seemed slightly bothered by my questions, and I was immediately overwhelmed by the prospects of having to eventually go back and make sense of all of the relevant details from our conversation - some of it very technical - that I was hurriedly scribbling in my notebook as he talked. But for all of the uneasiness that it provoked in me, this early interview turned out to be a turning point in the direction of my research.

As my conversation with Dr. Forrester unfolded during that rescheduled second encounter, he discussed the many different elements of the Consortium guideline program. This included a description of the *compendium*, a published list of all interventions recommended in the Consortium's guidelines, which health insurance companies frequently confer when determining which interventions to cover under their benefit plans. He went on to say that his interest in this type of tool was informed by the work he did as a clinical epidemiologist prior to becoming a physician, the branch of the wider field of epidemiology that applies statistical practices to improving clinical decision-making and care. Although he insisted on the centrality and importance of guidelines as a technology for guiding clinical care, Dr. Forrester also displayed what occurred to me as a curious ambivalence. As he phrased it, even the best guideline won't necessarily improve the quality of care -- especially for people of low socioeconomic standing -- and so 'figuring out why good care isn't being delivered is also an important task' (Interview notes). It was here that he emphasized his interest in something called *comparative effectiveness research*, which he explained as being 'the logical direction to go in.' In his formulation, since it is guidelines that inform the compendium, and since the compendium informs what kind of medical care is covered by insurers, 'looking at how different treatments match up to one another should be the obvious ultimate interest.' .

As comparative effectiveness was not something I had encountered up to that point in my research, the interview concluded with Dr. Forrester's suggestion that I look at a paper by Robert Brook (2009) that had recently been published in the *Journal of the American Medical Association*.

~

This brief commentary, entitled 'Possible outcomes of comparative effectiveness research,' focuses on the relationship between two key pieces of legislation that were coming to the fore in

around 2009. First, in February of 2009, President Barack Obama signed the American Recovery and Reinvestment Act into law (colloquially referred to as ARRA, or the 'Stimulus Package'). This was a multi-billion dollar funding bill that sought to reinvigorate the American economy in the midst of the Great Recession, touched off by the global financial crisis that had began two years prior. While a detailed accounting of these funds is beyond the purview of this introduction, one line item stands out as particularly germane to the discussion of evaluative research in biomedicine that runs throughout this thesis: among the US$155 billion of ARRA funds spent on healthcare, a total of $1.1 billion was apportioned to support the development of research and research infrastructure for conducting comparative effectiveness studies of healthcare interventions. Furthermore, it was also during this period that President Obama began proposing a series of reforms to the US healthcare system that would improve access to affordable healthcare for all Americans. This would ultimately take the shape of the Patient Protection and Affordable Care Act (typically shortened to the Affordable Care Act, PPACA, or simply ACA), signed into law in March 2010, and which authorized the establishment of an agency whose mandate was to coordinate comparative effectiveness research at the national scale (Hoffman et al. 2016).

Reflecting on the convergence of the ARRA and the ACA, Brook (2009) stresses the importance of understanding what types of practices will fall under the label of 'comparative effectiveness,' noting that '[d]iscussions to date suggest that most of the funds will be spent on comparing one clinical procedure, device, or drug with another' but also that there are 'a nearly infinite number' ways that such comparative studies can be conducted (ibid.:194). In his view, this infiniteness threatens to spiral the national CER program into a 'free-for-all,' or at the very least 'a full employment program for health services researchers and epidemiologists,' and it would be business as usual for the research design and application process: a researcher submits a proposal to conduct a given comparison between two drugs or procedures and an evaluation panel would evaluate the quality of the study based on the importance of the comparison being made, the feasibility of conducting the study, and the ability of the design to answer the specific research question being asked (ibid.). In principle, Brook finds 'nothing wrong' with conducting studies that test a new and more expensive treatment against existing standards of care to determine minor improvements in health. Yet against the backdrop of simultaneous efforts to improve the affordability of care for the

American population, and the insistence on rapid implementation of comparative effectiveness study results, he is led to 'question [...] whether the scarce resources allocated in the stimulus package should be used in this way' (ibid.).

We thus find the author emphasizing the need for 'an organizing principle to *guide the selection of which aspects of medical care to examine*' -- in particular, a framework that would require proposed comparative effectiveness research projects to satisfy two specific criteria prior to receiving any government funds:

> First, a grant or contract to spend public money must include an initial analysis to *establish a business case* that implementing whatever is being proposed would reduce the cost of care by a certain percentage. [...] Second, it would not be enough to establish that implementing the new drug, device, or other therapy will save money. The history of science shows that it takes a long time for new knowledge to be incorporated into day-to-day practice. So a second requirement [...] should be that successful innovations are implemented immediately. Thus, a successful application under the comparative effectiveness initiative must *include constituents, such as health care organizations, hospitals, physicians, or organized community groups,* that would agree to adopt the new therapy immediately if it were shown to be as safe as the old therapy but substantially less expensive (ibid., italics mine).

Brook warns that by adapting such a framework, federal agencies that fund CER might risk being 'accused of sponsoring rationing' (ibid.: 195). And yet, looking to the computer industry as a comparator, he argues that '[a] strong case can be made that this does not represent rationing' but is rather simply an instance of 'using [comparative effectiveness] research dollars to produce therapies that are both better and substantially less expensive' (ibid.). Nobody ever accuses the private sector of rationing computer chips when they are able to purchase a computer today at a fraction of what it would have cost in years past: 'Instead, the research and development model in the computer industry has been to make better machines, and to make them at increasingly lower costs, thereby making computers affordable to many more individuals' (ibid.). Analogously, he concludes that the time has come 'to use public funding and comparative effectiveness research to accomplish the same thing in medicine' (ibid.).

~

Brook's paper was far more prescient than I could have appreciated at the time of its initial publication, and there are several reasons why I have chosen to begin my empirical analysis with a somewhat lengthy rehearsing of his arguments. Of course, in the context of this opening narrative, it marks my very first point of contact with the topic of comparative effectiveness research, which would soon emerge as the primary object of my doctoral research. But it also presaged a series of distinct conceptual themes that would come to predominate in wider discussions around CER over the next several years. For one, comparative effectiveness is positioned as an approach to achieving *value* in healthcare: it proffers the use evaluative techniques of comparison and commensuration to guide the selection of the most effective biomedical interventions -- generally understood in terms of clinical utility, cost effectiveness, or some combination thereof -- and to rule out the use of those offering little value. This is linked to the theme of uncertainty around *defining* and *operationalizing* comparative effectiveness research in practice, at the very moment it first emerged in national policy contexts. Such uncertainty stood to hamper the likelihood of maximizing the societal value that CER promised in the first place. It is here that we find Brook's proposal for a framework that would require recipients of *public* money to involve *constituents* in comparative effectiveness research study proposals and to make a *business case* for -- that is, justifying on economic grounds -- the implementation of CER studies and their outcomes.

In what follows, I provide a conceptually-driven overview of the emergence of CER and its positioning within the wider ecology of evaluative research paradigms in biomedicine. I also engage in a more fine-grained ethnographic analysis of a single federally-funded comparative effectiveness research initiative -- the aforementioned Center for Comparative Effectiveness Research in Cancer Genomics -- whose work was largely structured by and oriented towards the foregoing list of concerns. As we will see, comparative effectiveness research was enacted within CANCERGEN as a response to the convergence of several key issues that many in the publicly-funded oncology research community viewed as contributing to growing practical inefficiencies and epistemic fissures in cancer clinical research and patient care. At the center of CANCERGEN's particular enactment of comparative effectiveness research were a core set of evaluative questions: What qualifies as comparative effectiveness research? How should the allotment of scarce public resources be prioritized for sponsoring research? Which methods and tools are best suited for assessing the value

of biomedical technologies and of implementing comparative effectiveness research findings? Who counts as a 'constituent' and what is a legitimate instance of 'involvement'? And how do project organizers and constituents reflexively evaluate their respective roles and positioning within comparative effectiveness projects?

The next section of this introduction (Section 2) describes the environment within which CANCERGEN first emerged in order to contextualize the subsequent characterization of CANCERGEN itself. I provide an overview of the high level definitional work being carried out within national policy debates that were ongoing at the time the project was first funded by the National Cancer Institute circa 2009-2010 as well as within a related discussion about the emerging relationship between CER and the group of technologies collectively referred to as *personalized medicine*. Recounting the set of issues that CANCERGEN project organizers deemed especially urgent, it becomes clear why their application of CER focused specifically on cancer genomics, and on molecular diagnostic tools in particular. This includes a problematization of the biomedical regulatory landscape in the US, through which personalized medicine technologies have historically proliferated into routine clinical practice without being required to demonstrate evidence that they improve patient outcomes.

Related to this was growing concern around the publicly funded cancer clinical trials system, which was posited as the primary site where studies testing the comparative effectiveness of these tools would largely be carried out. However, circa 2009 this system was approaching a state of crisis, the result of significant budget cuts for conducting studies, diminishing patient enrollment numbers in these studies, and an alarmingly high rate of studies being closed prior to their completion. This confluence of factors was viewed as contributing to wasted opportunities for assessing diagnostic tools as well as wasted government expenditures, which begged for methods to more efficiently prioritize and design CER studies.

From there, I move on in Section 3 to provide a review of the relevant literatures on valuation practices in sociological and STS scholarship in order to situate my own analytic approach. This culminates in my introducing Thévenot's *regimes of engagement* framework, which proposes its own unique take on (e)valuation understood in terms of judgments and testing of 'appropriate action' between actors and their environments in situations of coordination (Thévenot 2014b:246; cf.

Thévenot 1990). In particular, I outline the importance of Thévenot's notion of *engaging in plans*, the primary conceptual hinge for characterizing CANCERGEN's prioritization dispositif upon which much of the subsequent analysis is based. Section 4 concludes this introduction with a brief description of the remaining chapters and their main contributions, whereupon I segue in Chapter 2 to a presentation of materials and methods.

2. Situating the analysis

2.1 Defining comparative effectiveness research in policy

As we have just seen, it was in the lead-up to and passing of the American Recovery and Reinvestment Act in 2009 that comparative effectiveness research became a matter of national concern within health policy circles as well as among practicing researchers and clinicians. As a novel type of research with little legislative precedence, however, there was some inevitable uncertainty around what comparative effectiveness research was understood to be, both definionally and substantively.

On the policy side, several definitions would eventually surface, of which I discuss two frequently cited examples here (but see Section 4.2 and Chapter 3 for a more in-depth discussion of this topic). The very earliest of these was put forth by the ARRA-mandated Federal Coordinating Council for Comparative Effectiveness Research (FCC-CER), whose role would be to aid the coordination of federally funded CER studies as well as to serve in an advisory role in managing the allocation of the $400 million of CER funding dispensed by the Secretary of HHS (Birnbaum and Slutsky 2010). Its first charge, however, was to develop a working definition of CER, to which it responded with the following:

> Comparative effectiveness research is the conduct and synthesis of systematic research comparing different interventions and strategies to prevent, diagnose, treat and monitor health conditions. The purpose of this research is to inform patients, providers, and decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances.
>
> - To provide this information, comparative effectiveness research must assess a comprehensive array of health-related outcomes for diverse patient populations.

- Defined interventions compared may include medications, procedures, medical and assistive devices and technologies, behavioral change strategies, and delivery system interventions.
- This research necessitates the development, expansion, and use of a variety of data sources and methods to assess comparative effectiveness (FCC-CER 2009:5).

Congress subsequently tasked the Institute of Medicine (IOM) -- one of the US National Academies of Science -- with identifying national priorities for conducting comparative effectiveness research. As it worked towards enumerating these priorities, committee members looked to six extant definitions that had been proposed by other federal agencies and IOM committees to develop its own synthetic definition of CER:

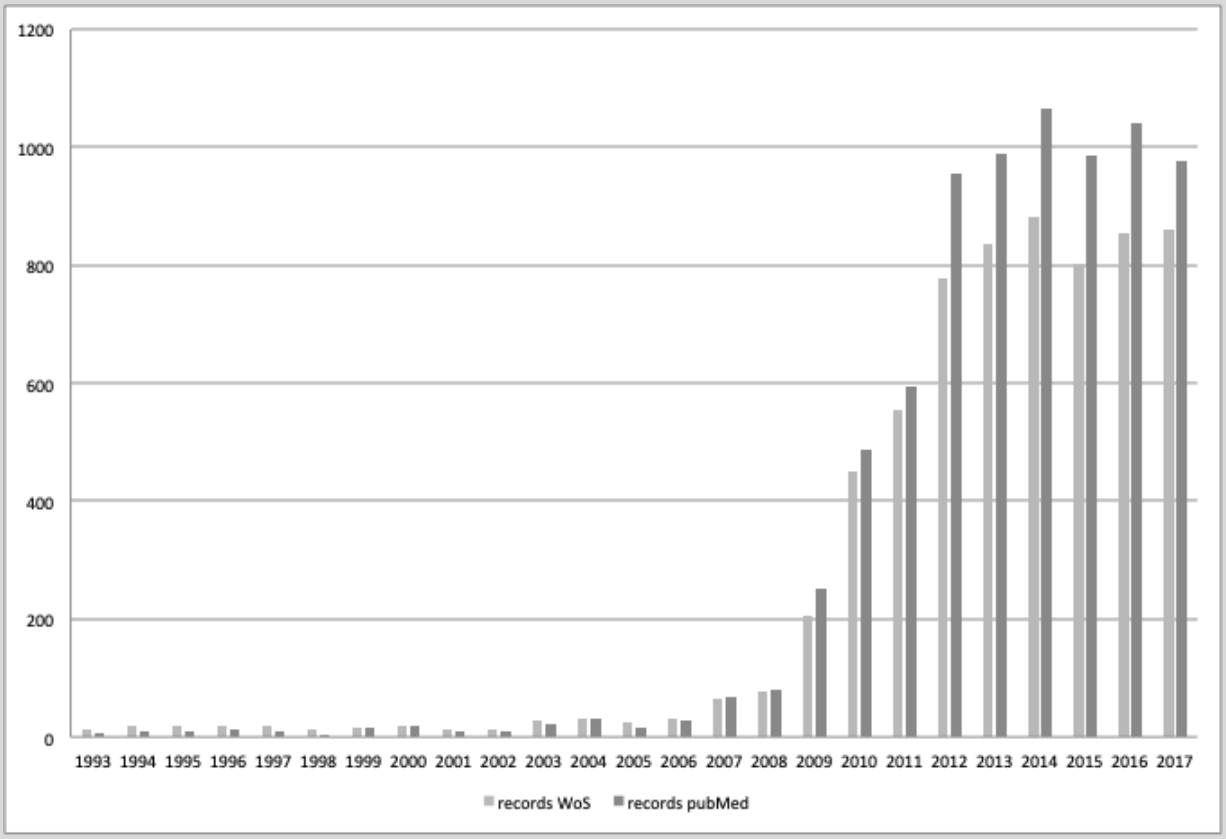
> Comparative effectiveness research (CER) is the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels (IOM 2009:13).

'These words taken together,' wrote prominent Institute of Medicine affiliate Harold Sox, 'define a form of research that would constitute a unique mission for a well-funded national program' (2009:S7).

Meanwhile, as this definitional work was being carried out at the level of national policy, there was a proliferation of more localized enactments where comparative effectiveness was being defined *in practice*. In these spaces, the contours of CER appeared to be much fuzzier than more established evaluative research paradigms in biomedicine such as health technology assessment, health services research, and evidence-based medicine (cf. Luce et al. 2010). What were pointed to as *instances* of comparative effectiveness research often incorporated many of the same practices of evidence generation and synthesis found in these other research practices. It is perhaps unsurprising to find Levy and Garrison (2010) observing at the time that 'there is still much debate on the essential elements of this concept' (S1). As an emergent type of research in its own right, we might phrase this in terms of the *interpretive flexibility* of comparative effectiveness, a concept which points to the

multiple future possibilities contained in a sociotechnical phenomenon during its earliest stages of development (Bijker et al. 1987).

Figure 1.1 Annual citation count of papers using the combined terms 'comparative effectiveness,' Web of Science and pubMed, 1993-2017



Looking to the literature, the rapid uptake of the combined term 'comparative effectiveness' leaves the reader with little doubt that the seeds for a national program were being sown at the time. Figure 1.1 shows annual citations for *comparative effectiveness* over a 25 year period (1993-2017) in both Web of Science, a cross-disciplinary citation indexing service, as well as pubMed, a citation index specific to biomedicine and the life sciences. As is clear from this graphical depiction, beginning in 1993 and lasting for over a decade, there were well under 200 papers published each year using the term *comparative effectiveness*. This pattern only began to change around 2006, which began a period of consistent year-over-year increases in term's use. Importantly, one notices a significant uptick in 2009 -- which we will recall was the year in which the ARRA apportioned over US$1 billion for CER --

and a doubling in annual counts during the 2008-2009 and 2009-2010 periods, moving from ~100 to ~200 and ~200 to over 400, respectively. Annual citations continued growing until 2013, after which they steadily reached a minimum of 800 yearly references, a trend that continues through to the end of 2017.

2.2 Researching and regulating personalized medicine

At the same time that CER was becoming a going concern in the US, so too did this period witness growing interest in and uptake of *personalized medicine* (PM). An accompanying set of concerns pertaining to how these technologies should be evaluated, regulated, and deployed in clinical practice (cf. Hedgecoe 2004, 2008) added further nuance to the interpretive flexibility of CER.

One issue was that personalized medicine could be defined in several different ways. Garber and Tunis (2009), in a *New England Journal of Medicine* discussing the CER-PM nexus, list off a few competing qualifications of personalized medicine including (1) a consideration of patient demographic characteristics such as age, comorbidities, and individual preferences; (2) the development of new therapies tailored to individual patients, such as vaccines and monoclonal antibodies; and (3) the implementation of genomic technologies that inform diagnosis, prognosis, and treatment based on a patient's genetic profile and other similar biomarkers. The latter qualification was especially prevalent in the field of oncology, where its promises were beginning to bear fruit in the mid-late 2000s. In breast cancer alone, examples included HER2 expression testing for predicting patients' response to the monoclonal antibody trastuzumab; OncotypeDX and Mammaprint, which are used for predicting patient benefit from toxic chemotherapy and for estimating the likelihood of disease recurrence following treatment; as well as BRCA1/2 testing, which can be used in guiding decisions about prophylactic surgery for those who manifest mutations of those genes (Bourret 2005; Kohli-Laven et al. 2010). For patients with certain lung and colorectal cancers, mutation testing for the KRAS gene was similarly beginning to inform treatment options, while testing for over-expression of epidermal growth factor receptors (EGFR) was being driven by the availability of targeted therapies that inhibit the receptors, thus slowing disease progression and metastasis (Normanno et al. 2009).

Second, there were also debates about the wider commensurability of CER and personalized medicine, and thus what role the former could play in evaluating the latter. Some went as far as claiming that the two domains were in fact mutually exclusive. This stems from an understanding of CER as operating, on the one hand, at the level of *universal patient groups*; CER studies operating according to this logic would essentially compare the effectiveness of different interventions for a single indication across these different groups. The corollary of this was the perception that personalized medicine operated at the level of the *individual patient*, and thus aggregate evaluations of effectiveness were simply not possible. By way of a rejoinder to such an understanding, others argued that CER may be alternatively conceptualized as the perfect tool for demonstrating whether personalized medicine technologies were capable of producing better patient outcomes than existing standard of care interventions. Garber and Tunis in fact argue that the it is not comparative effectiveness that poses the greatest challenge to implementing personalized medicine at ever-greater scales but rather 'the lack of adequately designed studies assessing [...] clinical utility' of these technologies, as well as the overall lack of 'consensus about the best way to design and implement such studies' (2009:1926).

Khoury (2009) framed this problem as the 'evidence dilemma' in the United States:

> [T]here is big emphasis on genomics discovery research but much less on research that allows discoveries to be evaluated for integration into practice, and for documenting their health impact among all segments of the population […] Even when we have good evidence for gene-based therapeutics or diagnostics, research on implementation, diffusion, and dissemination is often not done, and adoption in practice is uneven, under-resourced and not well distributed […] Right now, there may be a few genomic applications that are 'lost in translation.' However, currently a more prominent genomics translation challenge is 'premature translation' where such applications are not ready to be integrated into practice' (159).

But is there perhaps more to the story of 'premature translation' than simply a lack of consensus about trial design among the community of biomedical researchers operating in this space? Implied in Khoury's observation is a point about the ways in which personalized medicine technologies have -- and haven't -- been regulated in the US. It is important to specify here that the technologies of concern to these actors were not pharmaceuticals, which are without exception regulated by the US Food and Drug Administration based on the claims drugmakers make about their products vis-a-vis

safety and *effectiveness* (Marks 2009). Rather, it was the corpus of emerging personalized *diagnostics* technologies whose regulation seemed to be precipitating this dilemma. Ostensibly diagnostics regulation could, alongside drugs, fall to the hands of the FDA; historically, this has in fact not been the case however. Hogarth (2015) attributes the absence of FDA involvement in regulating certain diagnostic technologies to a wider historical shift in the agency's political-economic vision, noting a progressive retreat from its traditional role as gatekeeper of market access for these technologies. He frames this as a transition towards a collaborative mode of 'network governance' whose objectives are to facilitate innovation and more rapidly translate scientific discoveries into clinical practice through collaborative co-shaping of policy by private and public entities (4).

Beyond this more institutional explanation, issues of professional jurisdiction, technical complexity, and regulatory categorization have also come to bear on the regulation of these technologies, especially in the case of molecular diagnostics. For one, molecular diagnostic assays have typically fallen into the category of *laboratory-developed tests* (LDTs), which means they are processed at a central laboratory. Rather than circulating as *test kits* -- which would place them squarely within the FDA's regulatory purview -- laboratories that processes these assays have been considered to be delivering a *service*. As such, under a hands-off policy referred to as *enforcement discretion*, the FDA has ceded regulation to the Centers for Medicare and Medicaid Services, which oversees LDT regulation under the Clinical Laboratory Improvement Amendments (CMS/CLIA) (Evans et al. 2015). The evidentiary bar for laboratory certification according to this regime requires only that labs follow standard operating procedures and that the processed assays accurately measure the specific biological phenomena they claim to be measuring -- what is commonly referred to as their *analytical validity* (Bourret et al. 2011). As Sturdy's recent history of diagnostic regulation argues, the 'lack of attention to the clinical consequences of diagnostic tests' implied by this regulatory situation was a deliberate element of the FDA's strategy: 'Congress had repeatedly declared that the legislation authorising FDA to regulate medical products did not provide a basis for regulating medical practice. FDA interpreted this to mean that they should not interfere in how doctors chose to use diagnostic devices' (2018, under review).

2.3 The trials of publicly-funded CER in oncology

The regulatory transposition from the FDA to CMS/CLIA has meant scant legal and financial incentives for diagnostic manufacturers to fund research into many of the other technical properties of their products beyond what is required by the latter. This is to say nothing of the fact that, at the time CER became a going concern in the US, there existed no precedent for *any* federal agency to regulate market access of diagnostic assays based on a standard of improved clinical decision making or better health outcomes. In the oncology setting, this meant that evaluations of the clinical usefulness -- or what is referred to as the *clinical utility* -- of diagnostics fell outside the paradigm of formalized 'approval regulation' manifested by the FDA (Carpenter 2010). It thus largely fell to a host of other entities, including publicly funded research infrastructures like the National Cancer Institute's clinical trials Cooperative Group program (CGP). As we will see below, the CGP would come to play an important role in the Center for Comparative Effectiveness Research in Cancer Genomics and its particular deployment of CER; some background about this organization is thus in order here before moving on to describe how exactly it fit within the wider CANCERGEN consortium.

The Cooperative Group program first emerged during the 1950s, an innovative organizational structure in its own right that functioned as the longstanding 'platform for the design of new projects' within the NCI (Keating & Cambrosio 2012:86). Within its approach to planning, trialists were always already enmeshed in networks that iteratively drew on previous trials in designing and executing subsequent ones. As the program 'did not flow from a single model, nor did it emerge fully formed' in the postwar period, Keating and Cambrosio conceptualize the CGP in terms of pluralized 'cooperative approach*es*' (ibid.). Importantly, the program's legacy configuration was developed in explicit '*opposition* to a centralized, industrial model of organization' (ibid., italics in original). In March 2014, however, the NCI announced a significant transformation to publicly funded cancer clinical trials program. Replacing the existing multi-disciplinary and multi-institution CGP would be the National Clinical Trials Network (NCTN) which NCI CTEP Associate Director Dr. Jeff Abrams has described as a novel 'state-of-the-art clinical trials infrastructure' that is 'poised to implement and complete trials far more rapidly than in the past' and is stated as having special 'appeal for industry partners' (NCI 2016).

What is clear here is that the emerging role of personalized medicine was front and center in the NCI's public justification for its re-architecturing project. As the Institute's website states, knowledge in the area of cancer genomics has 'fundamentally changed our approach to cancer treatment,' which includes the deployment of targeted therapies driven by 'molecular methods [...] that will deliver optimum results' (ibid). Actualizing this 'form of precision medicine' is no easy feat, requiring the 'screen[ing] [of] large numbers of patients [...] to identify those patients whose tumors contain the distinct molecular targets of the therapies being tested' (ibid.). The NCTN was developed with these challenges in mind, 'organized to take maximal [*sic*] advantage of the opportunities afforded by the improved understanding of tumor biology as well as the improved efficiencies created by the *centralization and streamlining* of many [of the Network's] critical functions' (ibid., italics mine). The appeal for industry collaboration, moreover, comes through the Network's capacity to facilitate and streamline collaborators in 'harness[ing] next-generation DNA and RNA sequencing methods to inform treatment choices' for patients (ibid.). While the NCTN and its decidedly more centralized and industrialized vision certainly presented a remarkable departure from the decentralized and non-industrial organizational origins of its predecessor, the sunsetting of the CGP came only after a several year period during which the NCI conducted 'extensive consultation and coordination with many stakeholders' in the clinical research community (ibid.).

As this process unfolded, so too did numerous problematizations of the inefficiencies and inadequacies of the existing Cooperative Group structure, made manifest in the convening of expert panels and in the publication of a number of formal reports and commentaries (e.g. Califf 2009; Dilts 2010; Ramsey & Scoggins 2009; Twombly 2009). However, it is a 2010 Institute of Medicine report, jointly authored by the Committee on Cancer Clinical Trials and the NCI Cooperative Group Program Board, which stands as perhaps the single most important indictment of the CGP during this period (Nass et al. 2010). Few words are wasted in the preface of the document -- entitled, 'A National Cancer Clinical Trials System for the 21st Century: Reinvigorating the NCI Cooperative Group Program' -- before arriving at its central claim. The authors boldly state that 'the system for conducting cancer clinical trials in the United States is approaching a *state of crisis*' and '[c]hanges are urgently needed if we are to continue to make progress against the second leading cause of death in this country' (ibid.: ix, italics mine). It goes on to claim that the 'clinical trials infrastructure [...] has

not evolved to accommodate the rapid pace of biomedical discovery' and enumerates a series of challenges facing the CGP:

> Stagnant funding, inefficient processes, extensive and complex government oversight, and a growing trend toward the conduct of industry trials overseas have contributed to inadequate physician and patient participation in clinical trials, threatening the Cooperative Group Program's ability to efficiently translate discoveries into clinical applications (ibid.:64).

Among the 'urgently needed' changes are improvements of the 'efficiency' and the 'effectiveness' of the system itself; the report thus calls for the Institute to '*streamline* oversight, enhance collaboration, select and prioritize trials more stringently, fully fund the most innovative and promising studies, and open and complete trials with greater speed' (ibid.: xi). Left unaddressed, the already 'terrible waste of human and financial resources' would serve only to worsen the state of crisis in which the Cooperative Group program found itself (ibid.:x).

2.4 The emergence of CANCERGEN

It should by now be clear that, circa 2009, there was no shortage of moving parts within the comparative effectiveness research landscape in the US. At the level of national policy we find attempts to define comparative effectiveness research for the purposes of establishing a national CER program, while in the wider biomedical research community there surfaced a series of connected debates about its relationship to other modes of biomedical evaluation on the one hand, and to the emerging field of personalized medicine on the other. It is in the shadows of these three discussions that we may situate the Center for Comparative Effectiveness Research in Cancer Genomics, an organization whose work was motivated by several relevant questions: What are the defining features of comparative effectiveness research? In the absence of more formal regulation, could CER be deployed to evaluate the clinical utility of personalized medicine technologies? If so, how could it mobilize the community to reach consensus on how best to design and carry out those evaluations? And could the publicly funded NCI cancer clinical trials infrastructure somehow be both leveraged in this process, and perhaps also rescued from its descent into crisis in the same move? Drawing from a diverse cache of socio-organizational and socio-technical resources, CANCERGEN set out to provide a few answers.

Although billed as a *center*, the team of CANCERGEN project organizers was in fact composed of individuals from four distinct and geographically distributed organizational entities. This included a large academic cancer research center located in the Pacific Northwest of the US; a health outcomes research department of a large state university, also located in the Northwest; a Mid Atlantic-based nonprofit organization specializing in practices of healthcare stakeholder engagement; and the (paradoxically located) Michigan-based SWOG (formerly the Southwest Oncology Group), one of what were at the time ten NCI clinical trials Cooperative Groups. In addition to this diverse group of participating institutions, there was also a fifth entity -- the 13-member External Stakeholder Advisory group (ESAG) -- whose members were invited to participate in the project as representatives of various decision-making constituencies within the American healthcare system. Among this group were patient and consumer advocates; local, state, and national policymakers; payers (e.g. insurers); practicing clinicians and clinical researchers; as well as individuals from the medical device industry. Project organizers envisioned CANCERGEN as a 'proof of principle' platform for prioritizing and designing large-scale comparative effectiveness trials of diagnostic technologies.

This spoke directly to a concern expressed in the aforementioned IOM report, which highlighted that one of its 'major challenge[s]' was 'the prioritization and selection of trial concepts before a trial is launched. The effective prioritization and selection of trial concepts is critical to ensure that limited public funds are used in ways that are likely to have the greatest impact on patient care' (ibid.:29). Implementing this vision entailed the deployment of a heterogeneous ensemble of evaluative techniques (described in further detail in Section 4.5 below and later in Chapter 6). For instance, during its earliest phases, project organizers conducted a *landscape analysis* of the many personalized medicine technologies -- primarily diagnostic assays -- that were either coming down the development pipeline or already in routine clinical use (Thariani et al. 2012). This involved conducting an evaluation and formalized review of the biomedical literature, as well as consultations with various clinical experts in the fields of oncology and personalized medicine, with an eye towards isolating a shortlist of tests that were deemed ready to submit to the rigors of a CER study, and which would then carry forward into the next phase of the prioritization and design process. A subsequent phase of evaluation involved the aforementioned ESAG, whose members were tasked

with further prioritizing the six diagnostic technologies which were identified as the top study candidates in the landscape analysis.

This second phase was itself parsed into several rounds of stakeholder deliberations, evaluation, and voting, during which time a still further socio-technical intervention was introduced: a hybrid economic-decision modeling method called value of information analysis (VOI, discussed in Section 4.4 below and later in Chapter 5). Using this modeling approach, project organizers sought to introduce novel forms of quantitative, economized evidence into the stakeholders' evaluations of the six tests with the aim of arriving at a still further shortlist made up of the three highest-ranked tests of the initial six. This, in turn, would come to inform the final phase of evaluation, where it was expected that the stakeholders would reach a consensus on a single top-ranked diagnostic technology with which to carry forward into the design of a large-scale comparative effectiveness study. Still further implementations of the VOI modeling intervention were deployed during this design phase, which was to culminate in carrying out of the trial within SWOG's Cooperative Group infrastructure. Value of information analysis was also proposed as part of a longer-term vision for the Cooperative Group program, which could possibly take up VOI on a larger scale to streamline the oversight of the clinical study approval process, more efficiently prioritize and design those studies, and execute them with greater speed and certainty about their completion. In this regard, CANCERGEN was to be the first of several experiments in testing the fit-for-purposeness of VOI.

Before moving on to an examination of the execution and implications of these various evaluative plans in our empirical chapters, however, for now I turn our attention to a discussion of the relevant (e)valuation literatures in the fields of sociology and science & technology studies.

3. Reviewing the literature: (e)valuation as theory and practice

3.1. Disciplinary boundaries in the study of values and evaluation

Scholarship in the fields of sociology and science & technology studies has over the past several years taken a renewed interest in themes of value and (e)valuation. Of course, one can convincingly argue that such an interest has always been at the heart of sociological research, and perhaps at the very core of the social sciences writ large -- inspired as they have been by Marx's analysis of capitalism, Weber's work on legitimation, and Durkheim's influence on a moral and juridical sociology, and extending into the mid- and late-20th century works of Parsons and Bourdieu on

rationality and symbolic capital, respectively (Cefaï et al. 2015; Lamont 2012). But as Michèle Lamont (2012) has recently argued, although the Sociology of Valuation and Evaluation (SVE) is not a 'recognized subfield,' on both sides of the Atlantic there has been a 'remarkable' upsurge in sociological works 'concerned with how value is produced, diffused, assessed, and institutionalized across a range of settings' (203).

For her, this is indicative of both a broader interest in studying fundamental social processes -- including standardization (e.g. Bowker & Star 1999; Timmermans & Epstein 2010), boundary work (e.g. Gieryn 1999), and commensuration (Centemeri 2015a; Espeland & Stevens 1998, 2008; Fourcade 2011a, 2011b) -- and the preoccupations of those who deploy a *social problems* framing: growing meritocracy, societal inequality, and the proliferation of new tools of performance evaluation that reach across societal sectors and national jurisdictions. Justifying the centrality of (e)valuation as a topic for sociological analysis, Lamont qualifies it as a distinctly socio-cultural process:

> establishing value generally requires (*a*) intersubjective agreement/disagreement on a matrix or set of referents against which the entity (a good, a reputation, an artistic achievement, etc.) is compared, (*b*) negotiation about criteria and about who is a legitimate judge [often involving conflicts and power struggle] [...], and (*c*) establishing value in a relational (or indexical) process involving distinguishing and comparing entities (2012:205).

In this regard, a sociological approach to (e)valuation is equipped to account for an altogether different set of phenomena than what is otherwise found in the cognitivist perspectives of psychology or the behaviorist perspectives of economics. Yet despite this proliferation, others have argued that instead of working to settle the contours of a *sociology of (e)valuation* as a specific sub-discipline of the wider sociological field, perhaps we would do better to approach it as a '*focus of perspective*, transversal to all the social sciences as soon as they come to terms with the ways actors assign meaning to what they do, and to their transactions with things and people' (Cefaï et al. 2015: 2). This argument is reflective of a tension initiated in the mid-20th century that continues to persist in the present era, namely around the negotiation of disciplinary boundaries and the division of academic labor in the study of value(s) and (e)valuation.

As David Stark (2009) recounts, it was over 60 years ago that the American sociologist Talcott Parsons, based at Harvard University, was working over his 'grand designs' of reshaping the social sciences. Although he had confidence that in his 'imperial' ambitions he could lay claim to the fields of sociology, psychology, and anthropology, he was less certain about the discipline of economics, which he perceived as possessing the ability to 'thwart his agenda if his program was perceived as encroaching on its territory' (ibid.:7). And so he walked down the hallway and conferred with his colleagues in the department of economics, where he told them of his plans and 'assur[ed] them that he had no designs on their terrain' (ibid.). 'Thus,' Stark concludes, 'Parsons made a pact' which essentially said: 'You, economists, study *value*; we, the sociologists, will study *values*. You will have claim on the economy; we will stake our claim on the social relations in which economics are embedded' (ibid., italics mine). Parsons' Pact thus sets up a dichotomy, and confronts the social scientist with a choice between studying the singular (economic) value and the economy on the one hand, or the plural (social) values and societal relations on the other. Stark, however, argues for the need to move beyond such simple bifurcations, using the 'polysemic' notion of *worth*, which 'signals concern with fundamental problems of value while recognizing that all economies have a moral component. Rather than the static fixtures of value and values, it focuses instead on ongoing processes of *valuation*' (ibid.:7-8, emphasis in original).

3.2 Pragmatist and pragmatic approaches to valuation

We will return to the notion of *worth* shortly. For the moment, though, the important element in this retelling of the genesis of a subfield of economic sociology -- and its jurisdictional differentiation (Abbott 1988) from the discipline of economics -- is the shift in analytical emphasis from *value(s)* to *valuation*. Equally clear in Lamont's aforementioned description of (e)valuation as in Stark's own, this turn to *valuation* is considers values only in the context of *processes of value-attribution*, a point notably made by the American pragmatist philosopher John Dewey (e.g. 1939) that has been taken up as a common reference in more recent work on the topic. Muniesa (2012), for instance, points to Dewey's *flank movement*, a term the latter used when responding to an ongoing early-20th century philosophical debates between two groups. The 'idealists,' on the one hand, held that value becomes attached to things independent of human consciousness; those of a more 'realistic' orientation, on the other, argued for the necessity of consciousness as mediating the value of things.

Flanking the terms of this debate, Dewey instead argued for a shift in grammatical form from value to valuation 'considered explicitly as an action':

> [s]peaking literally, there *are* no such things as values [...] There are things, all sorts of things, having the unique, the experienced, but undefinable, quality of value. Values in the plural, or value in the singular, is merely a convenient abbreviation for an object, event, situation, *res*, possessing the quality. Calling the thing a value is like calling the ball struck in baseball, a hit or a foul (Dewey 1923, quoted in ibid.:25).

Dewey's later writings on valuation show him doubting such a formulation, however, as when he states that 'privileg[ing] [...] the abstract noun in the analysis of value generally gives rise to confusions and ambiguities' (Dewey 1949 quoted in Quéré 2015:205). Such a consideration leads him further to emphasize the *adjectival* property of value: it 'nam[es] that which is a trait, property, qualification of something [...]. It is like, say, the words good, fine, excellent. [...] If language had provided us with a special abstract noun (such as goodness in connection with good), say valuity or valueness, a good deal of ambiguous discussion resulting in incoherent conclusions might have been avoided' (ibid.).

Dewey's various grammatical formulations notwithstanding, the abiding point is how the flank movement positions valuation as a process and, perhaps more importantly, as a *practical* activity. Neither a mental operation nor a subjective practice, 'the situation in which judgment of value is required [...] exists as something whose good of value resides (first) in something to be attained in action and (secondly) whose value both as an idea and as existence depends upon what to do' (Dewey 1915 quoted in Muniesa 2012:25-6). 'Value,' he concludes, 'is "objective," but it is such in an active or practical situation, not apart from it' (ibid.). Understood as a form of action, then, Dewey's definition of valuation furthermore considers that it is empirically observable: 'The test of the existence of a valuation and the nature of the latter is actual behaviour as that is subject to observation' (Dewey 1939 quoted in Quéré 2015:166). This has important consequences for the social scientific analyses of valuation since adhering to the practice-based pragmatist understanding outlined above requires the analyst to avoid essentializing the notion of *value*.

In line with these precepts, Dussauge and colleagues (2015) have proposed a *valuographic* research program that 'encourages us to examine how certain things come to be considered valuable and

desirable, as well as how certain registers of value are ordered and displaced' (268). Approached in this way, valuography 'makes it possible to take an interest in values while moving away from from the question of what values "really" are' (ibid.). Stark's own turn towards a lexicon of *worth* is instructive here as well, as it presents one particular approach to considering the multiplicity, and indeed the *heterarchy*, of evaluative principles put into play in moments of valuation: 'Just as post-Mertonian studies of science moved from studying the institutions in which scientists were embedded to analyzing the actual practices of scientists in the laboratory,' he says, so too can 'economic sociology move from studying the institutions in which economic activity is embedded to analyzing the actual evaluative and calculative practices of actors at work' (Stark 2009:10).

Here, he draws from the sociological theory of value presented in the collaborative work of Luc Boltanski and Laurent Thévenot, most notably their book entitled, *On Justification: Economies of Worth* (2006). Working in an area of sociology referred to as *pragmatic sociology* or the *convention school*, these authors have developed a sophisticated 'horizontal plurality' of evaluative principles wherein they outline six distinct *orders of worth*: domestic, industrial, fame, inspiration, market, and civic. The authors trace each order (in French, this is translated as either *cité* [city] or *monde* [world]), representing a distinct vision of the *common good*, to a foundational text in political philosophy.

In critical moments of public dispute and contestation, actors draw on these orders when justifying themselves, pitting one version of the common good against another; they are, in Stark's words, 'distinctive and incommensurable' (2009:12; cf. Centemeri 2015a). This approach takes seriously the critical capacities of actors themselves who possess the ability to make such justifications, inverting the traditions of 'critical sociology' to instead perform a *sociology of critical capacity* (Boltanski & Thévenot 1999). Accordingly, in these critical moments people and things are subjected to *reality tests*, which

> leads the persons involved to agree on the relative importance of the beings that turn out to be implicated in the situation, whether the issue is, for example, the relative usefulness of two machines or two investments, the relative merits of two students, the competence of two business executives, or the tokens of respect that two local dignitaries owe one another. Very diverse beings - persons, institutions, tools, machines, rule-governed arrangements, methods of payment, acronyms and names, and so forth - turn out to be connected and arranged in relation to one another in groupings that are

> sufficiently coherent for their involvement to be judged effective, for the expected processes to be carried out, and for the situations to unfold correctly. [...] In order for the system to be open to judgment with reference to a higher common principle, each being (person or thing) has to be adjusted to it. When these conditions are fulfilled, we can say that the situation "holds together" (Boltanski & Théveot 2006:40-41).

In other words, what Boltanski and Thévenot describe as reality tests constitutes a type of *qualification trial* in which qualities of entities are 'attributed, stabilized, objectified and arranged' (Callon et al. 2002: 199). In so doing these, otherwise *particularized* entities undergo a process of *generalization* (*montée en généralité*) according to their correspondence to a given order. It is in this sense that we can say that these evaluations are thus at once *moral* and *practical* (Boltanski & Thévenot 2006:41).

Importantly, as Stark points out, a key insight offered by this approach is that the different orders of worth are each constitutive of value rather than counterposed to it (Stark 2009:11). Orders of worth are, moreover, said to produce the conditions of possibility for coordination in an uncertain world: they become evaluative conventions upon which participants in a given situation can agree and, in doing so, provided shared evaluative criteria on which actors may draw in their judgments and evaluations of what people, things, and actions are legitimate. Counterposing one order of worth against another, however, constitutes a type of friction -- something that Boltanski and Thévenot appear to frame almost exclusively in negative terms, i.e. that it is only through rare and costly *compromises* between multiple competing orders that actors can avoid the domination of one order over others in critical situations.

Stark's subtle critique of the orders of worth approach, on the other hand, argues that the imperative for universal agreement forecloses the possibility of disparate understandings 'so as to forestall an agreement, resulting in a breakdown of coordination' (ibid.:195). In his own work, he has sought to show the affordances of a simultaneous interplay of competing evaluative principles, what he terms *heterarchy*. Surveying economic activities across a myriad of industrial sectors and geopolitical conditions, Stark has found that innovation is 'facilitated not by convergence or agreement on a principle of justification but [...] by the *collision* of evaluative principles' (Stark 2017:388, italics mine). In this regard, amidst such clashes, 'novel recombinations become thinkable'

and so organizations supporting this simultaneous interplay are able to create wealth (ibid.). Ann Westenholz (2012) has similarly questioned Boltanski and Thévenot's privileging of orders of worth in stabilizing coordination practices, arguing that coordination need not require agreement on a single common evaluative principle, nor even a mode of *ordering coordination* -- her term for achieving common compromise across multiple orders of worth. Drawing from an empirical study of open source software communities, she demonstrates the existence of a *disordering coordination*. This arises when actors' 'activities are coordinated through *similar actions without reference to a common world or a common compromise*. It takes place because people perform the same action but do not agree on the meaning of this particular action' (ibid.:120, italics in original).

3.3 Reconceptualizing valuation in *regimes of engagement*

An assessment of the direction that Boltanski and Thévenot's respective research programs have taken following the initial publication of *On Justification* reveals that the limitations of this earlier work have not escaped their own attention, either. Boltanski (e.g. 2011), for instance, has problematized the 'lack of critical impetus' present in his earlier work on the *sociology of critique* developed together with Thévenot and has since attempted to reintegrate a more critical stance (Hansen 2016:136). Admitting notions of 'false consciousness' back into the analytic frame, this work has effectively also reintroduced the exteriority of the sociologist, precisely what the development of a sociology of critique sought to transcend. Summarizing this renewed state of affairs, Hansen (2016:37) observes that '[o]nce again the role of the researcher is "unmasking immanent contradictions" (Boltanski, 2011: 12), "identifying processes of exploitation [...] rooted in very unequal distribution of property" (p. 155) and finally "emancipating dominated classes" (p. 154)' (Hansen 2016:137). The sociologist is then well positioned to deploy her analytic toolbox and provide critical argumentations to the actors, who are assumed to possess a diminished critical capacity of their own (ibid.:138).

Thévenot, in his turn, has similarly problematized the orders of worth approach developed together with Boltanski; but while agreeing with certain tensions that Boltanski points to as justifying the need for a (re)turn to externalized critique, Thévenot's own recent research program stems from an altogether different view of the shortcomings encountered in *On Justification*. In particular, he argues that the 'horizontal pluralism' of orders of worth do not exhaust the totality of evaluative formats one finds in the world, and that the emphasis on justification and critique in the former

work tends to overshadow the costly operations of presenting arguments in formats that are *publicly* intelligible (Thévenot 2015b; cf. Centemeri 2010). Reflecting on this progression, he notes that the original analysis in *On Justification* was 'deliberately limited […] to those forms of commonizing that enjoy the greatest legitimacy, forms that channel uncertainty into coordination frames appropriate for public judgment and that imply a dynamic of critique and justification' (Thévenot 2007b:411). Acknowledging that '[i]nstitutions in action' involve a myriad of 'engagements with the world' (Thévenot 2006:4), Thévenot extends 'the analysis to action conceived as plural, seeking to handle the variety of cognitive and evaluative formats' that are less-well equipped for 'commoniz[ing] cognition to equal degrees' (2007b:411).

Whence Thévenot's articulation of *regimes of engagement* (ROE), a novel conceptual framework which seeks to provide a broader view of the tensions between commitment and doubt that are 'at the heart of all attempts to find guarantees or assurances' -- that is, not simply at the limited interface of institutions and language, as Boltanski would have it (Thévenot 2014b:249). Engagement, in its slightly idiosyncratic usage here, 'suggests the quest for a certain kind of insurance in the relation with the world, and draws attention to the correspondence between a capacity or power of the agent and the appropriate preparation of the environment' (Thévenot 2006:4.). In this way, his vocabulary departs from the more human-centric notions of 'action' and 'practice' in order to account for a material conceptualization of engagement with one's environment (cf. Marres 2012). But a second reason why he uses a vocabulary of engagements is to stress the 'quest for a guaranteed good' as in the case of contracts or marriages (Thévenot 2007b:415).

Engagements, in other words, equip actors with criteria for evaluating the 'relevant things' to be known, which are themselves equal to 'pledges that guarantee the good that fuels each regime' (ibid.). To wit:

> The *good* that engagement with the world aims to *guarantee* orients how reality is grasped and specifies the *format* of what constitutes *information*. An engagement lends itself to *communication* to varying scope depending on the format; the place and use of *language* also vary by format. It is from his dependence on an *engaged* environment that the *agent* derives his *capacity*, understood as the *power* to maintain that engagement. (ibid., italics in original).

Approached as a set of interlinked *situational dispositifs* (e.g. Dodier & Barbot 2016) -- sociomaterial assemblages that take form amidst a succession of instances and according to a given set of qualifications -- each regime has a particular conception of the *mutual engagement* in play; an *evaluative good* towards which action is oriented; a formatting of actors' *agency*; as well as a formatting of *information, objects*, and the *environment* in which they circulate (Table 1.1). Importantly, the ROE architecture is 'a comprehensive framework' and as such remains attentive 'to the role played by *evaluation* in shaping the dynamics of action in the world and delineating *regimes of valued engagement* between a human agent and her counterpart in the material environment' (Thévenot 2006:3, italics mine). It spans the most intimate regime of *familiar engagement* through to the most conventional *regime of public justification*, while also accounting for regimes of *engaging in exploration* and *engaging in plans*. In this way, it expands the discussion of public qualifications and critique outlined in the pages of *On Justification* and moves on to account for engagements 'far below this highest level of commonality' and which are 'far less prepared' for the type of 'public coordination' discussed above (2006:4).

Thévenot's approach has been taken up in a number of studies covering vast empirical ground, from controversies over access to dementia drugs (e.g. Moreira 2012) to studies of third-party certification in sustainable agriculture (e.g. Cheyns 2011) to international land governance frameworks (e.g. Silva-Castañeda 2016). Much of this work has, however, tended to focus on oscillations between the *familiar* and the *public* regimes, with valuation in *exploration* and *planning* remaining relatively overlooked vis-a-vis the wider framework (cf. Cochoy 2014; Kessous 2015).[1] Given the empirical material with which this thesis is concerned, one of its central contributions to demonstrate the affordances of the regimes of engagement architecture as well as to further develop elements of the *planning* regime that remain undertheorized in both primary and secondary literatures.

[1] This is perhaps more true of anglophone scholarship than what we find in France; indeed, much of the work that has fleshed out the conceptual core of the ROE framework has not been translated into English, e.g. Auray (2016) on the regime of exploration and Thévenot (1995) on the planning regime.

Table 1.1 The *regimes of engagement* dispositif: elements and qualifications ***Adapted from (Thévenot 2014a:135-135)***				
Regime >>> **Component:**	**Public**	**Plan**	**Familiar**	**Exploratory**
Mutual engagement	Legitimate convention of coordination	Joint project, contract	Close friendship, intimacy	Play
Agency	Qualified, worthy	Willful, autonomous stakeholder	Attached personality	Curious, explorer
Evaluative good	Qualifying for the common good, worth	Accomplishing stated plan of action	Feeling at ease	Excitement for novelty
Formatting of environment, information + objects	Conventional	Functional	Congenial	Surprising

3.4 The value of planned engagements

As with the other regimes, in planning it is the *engagement* itself which is the object of valuation, while the regime's dispositif in turn provides actors with a set of tools for evaluating the fitness of activities, of people, and of objects as they relate to the planned action, whatever that may be. To a certain extent, speaking of of 'engaging in plans' is but a prolix way of what might otherwise be described in terms of 'people just trying to get shit done.' But stating it so simply threatens gloss the specificity of this singular mode of 'doing things together' in the world (Boltanski & Thévenot 1999:359; cf. Becker 1986), which can be placed within a wider ecology of engagements. It also does not account for the significant investments that are required for formatting planned situations, the sacrifices that accompany those investments, and the myriad possible tensions that may arise when numerous actors or groups operating according to divergent frames of reference and understandings of what is *really* going on must nevertheless coordinate with one another to achieve the desired

outcome. With all but a few exceptions, theorizing on such ways of coordinating that we may refer to as *engaging in plans* has received scant attention in the social sciences due to the sheer functionality of discourses, objects, and human intentions one finds in such situations (cf. Agre & Chapman 1990; Eranti 2018; Flichy 1995; Suchman 1987; Tavory & Eliasoph 2013).

Lucy Suchman's (1987) foundational contributions to this literature, for instance, found through detailed observations of technicians and users of expert computer systems that plans are not simply matters of predictable execution or 'controlling structures,' but are rather *resources* that actors produce and use in their quotidian engagements. In this way, plans 'elaborate actions just to the level that elaboration is useful; they are vague with respect to the details of action precisely at the level at which it makes sense to forego abstract representation, and rely on the availability of a particular embodied response' (188). Citing Hutchins' example of Micronesian kayakers finding their way through the night in the absence of either drawn out plans or sophisticated navigation instruments, Suchman demonstrates how they 'maintain their course by substituting other environmental referents, that are accessible, for the inaccessible land' -- for instance by following a star path -- and that 'maintain[ing] their orientation to the star path at any given point in their voyage requires that they consult not only the stars, but a rich set of changing environmental circumstances [...] which through experience become interpretable information about the relative position of the canoe' (ibid.:187). From this, Suchman posits disjuncture between *plans* as ideal representations, on the one hand, and *situated action* as 'local interactions with the environment' on the other, nevertheless concluding that 'purposeful action' is the inevitable outcome of their mutual entwinement (ibid.).

More recently, Tavory and Eliasoph (2013) have singled out plans as a specific type of future-oriented coordinating device, noting their 'rarely dramatic or explicit' quality (916). Instead, plans tend to naturalize their very formatting; they occur to actors as 'almost automatic,' fostering an assumption that the only thing required is to follow the requisite steps of a path that has already been laid out (ibid.). While there is 'some leeway for individual volition' in plans, it is upon breakdown that actors typically become retrospectively aware of their 'intuitions' about what was expected of them in the 'ongoing activities' of carrying out the plan -- or 'how to be a good character in this story,' as it were (ibid.). Actors' successes or failures, then, are in fact preconditioned by earlier *investments in form*: 'After this initial setup,' the authors conclude, 'the actors treat the outline

of the future as if it were a natural part of their world, and only then [...] can they creatively improvise on the given theme' (ibid.:917-17; cf. Thévenot 1984).

Fitting at the nexus of these two aforementioned perspectives, Flichy (1995) approaches plans as a *frame of reference*. In doing so, he seems to agree with Suchman when he argues that plans do not '*determine* technical action in any way,' while at the same time gesturing to Tavory and Eliasoph in his assertion that plans offer 'a point of anchorage' that enable specific sociotechnical activities and in the very same move also impose a 'set of constraints' on situations (19, italics mine). In the case of the *planning* regime, actors' ideas and beliefs are formatted as individual opinions to be negotiated among a group of actors as they work towards executing a plan with a desired end in mind. In other words, plans establish the grounds for a particular *style* of coordinating that lays out how actors are expected to act vis-a-vis the regime's evaluative good (cf. Eliasoph & Lichterman 2003). While this may streamline decision-making, for instance, it also constrains acceptable forms of expression, precluding the recognizability of *familiar* concerns and gestures on the one hand, and demanding that actors not escalate to criticizing the wider purposes of the activity in a mode of *public justification* (Cheyns 2014; Silva-Castañeda 2012). This latter point suggests an important and generalized quality of regimes of engagement -- that they are inherently two-faced: on one side is the act of *closing one's eyes*, having *trust* and *confidence* in a given way of coordinating with oneself, with others, and with the material world; on the other, expressions of *doubt* and *inquietude* towards a particular mode, the act of *opening one's eyes* in an evaluation that considers how the situation *could be otherwise* (Blokker & Brighenti 2011; cf. Hughes 1984:552).

As can be gleaned from the preceding discussion, in acknowledging this vertical 'plurality of sources of confidence in our relation with the world,' social studies of (e)valuation open themselves up to 'a more sophisticated notion of the variety of senses of reality' which are limited neither to the *public* realm, nor to formal institutions, but include considerations of familiar engagements, of exploration, and of planning as we have outlined above (Blokker & Brighenti 2011:388). But in light of this review, I contend that Thévenot has himself has not provided the same level of conceptual explication vis-a-vis the regime of *planning* as one finds with the other regimes of engagement. For the purposes of the present thesis, then, I seek to push further the conceptualization of plans in a way that we are more readily equipped to confront modes of qualification and (e)valuation that do

not require the same investments as the most publicly legitimate forms, which in turn provides the analyst with a more granular analytic vocabulary for studying the social world. As we will see in the chapters that follow this leverage is critical to understanding how actors approach their engagements and how they go about evaluating and valorizing (Vatin 2013, cf. Centemeri 2015b) the people, things, and situations with which they are confronted as they move about in the world.

Homing in on orientations toward specific *goods* moreover enables an appreciation of value that goes beyond the tendency to parse the notion value into distinct registers or domains -- like 'ethical' value, 'social' value, 'economic' value -- by tying together networks of people and objects in discrete *situations*. Each situation, in turn, mobilizes distinct understandings of what and who gets to count, which at times are confronted by competing understandings within that self-same situation. This also enables novel forms of critical analysis that are focused not on personal dispositions and their role in propagating oppression, but rather on the situational formattings which require of actors different types of proof and are propitious to variable forms of commonality at different moments in time. By extension, it also renders available a mode of analyzing situations in which little or no critique actually occurs -- which of course does not mean that *nothing at all* is happening but merely that it is occurring in a different register of activity. The ROE architecture helps draw into relief the contours of that activity relative to other possible modes of engaging in the world, enabling the analyst to comprehend the sacrifices that are made when actors commit to a given framing of the situation, as well as to unpack the available resources that either enable or otherwise preclude one's ability to shift registers, e.g. to escalate from 'normal action' to a questioning of the underlying principles of that action.

4. Chapter overview

The chapters that follow address comparative effectiveness research in terms of an assemblage of *folded* (e)valuations, that is, as 'instances where one valuation practice is feeding and impinging on another valuation practice and the activities that achieve such interrelations between valuation practices' (Helgesson 2016:100). A consideration of such foldings, Helgesson argues, helps the analyst to engage with the 'complexity of interrelations and activities' between and within such

evaluative situations and practices that may be otherwise overlooked when attention is focused only on single instances.

4.1 Materials & methods

As an extension of this introduction, Chapter 2 provides an in-depth discussion of the empirical materials and methods of data collection and analysis upon which the ensuing arguments have been constructed. This includes a chronology of the shifting methodological toolbox I deployed over the course of my research, as well as a recounting of the grounded theory-informed process whereby intertwined questions of publicity and evaluation came to inform the central theoretical and conceptual contributions I make in the remaining empirical chapters.

4.2 Comparative effectiveness research in health technology assessment

Chapter 3 provides an overview of the development and deployment of comparative effectiveness over the past several years, with particular attention paid to its intersections with other modes of evaluation in health and biomedicine, especially that of health technology assessment (HTA). The relationship between CER and HTA can be characterized by something of a mutual folding yet the nature of the correspondence and relationships between CER and HTA as discrete evaluative practices was still very much an unsettled issue around the time that the former first emerged as a going concern. In the broader biomedical research and health policy literatures, individuals took up a range of positions with regard to the methodologies that should be mobilized in answering specific clinical questions and debated about the various standards for adjudicating evidence on the utility of biomedical interventions. Yet while much of this activity was geared towards a type of boundary work (Gieryn 1999), in which actors representing competing perspectives worked to parse and prescribe what CER and HTA *ought* to be, Chapter 3 of this thesis offers instead an historical–empirical approach to understanding the state of play of the HTA-CER nexus.

In so doing, it addresses two central themes introduced in our previous review of Brook's (2009) commentary: the advent of practices aimed at engaging healthcare stakeholders and communities in designing comparative effectiveness research studies, and the increasing attention paid to demonstrating the economic and clinical value of CER studies and their results. These concerns, however, are weaved into two further discussions: the first centered on the affordances of competing

research methods and study designs, including clinical trials, observational studies, systematic reviews, and meta-analyses; and the second on how CER and HTA could be deployed within a growing interest in patient-centeredness and personalized medicine. Following a review of these interlinked conversations, the chapter introduces the Center for Comparative Effectiveness Research in Cancer Genomics as a specific instantiation of CER. In this brief case study, I demonstrate some of the ways in which the debates occurring in the extant literature were deployed and addressed in the everyday practices of prioritizing and designing CER studies of diagnostic technologies in the field of oncology. Given the fact that its work was organized in such a way that addressed each of the four thematic axes just described, with a particular focus on the novelty of stakeholder engagement, I accordingly situate CANCERGEN as a paradigmatic example of comparative effectiveness research worthy of further exploration.

4.3 When participation goes according to plans

Diving deeper into the work of CANCERGEN, Chapter 4 takes as its focus the conceptualizations and practices of stakeholder engagement as they were folded into a wider project aiming to prioritize and design comparative effectiveness studies of personalized medicine technologies in cancer care. As we have seen throughout this introduction, engaging relevant actors and communities in these activities was a hallmark feature of the rollout of a national comparative effectiveness research program. But deciding what counts as a legitimate instance of 'engagement' and defining the relevance of specific actors and communities to be included in these processes as 'stakeholders' is an inherently political process with its own set of evaluative practices, judgments, and justifications. In the case of CANCERGEN, project organizers defined *stakeholder engagement* as something altogether different from *patient involvement* and *public participation*. Much of the sociological and STS literatures on participation in science and biomedicine has tended to focus on these latter two participatory formats, many of which -- including Nowotny et al.'s (2001) notion of the 'agora' and Callon et al.'s (2009) 'hybrid forum' -- display a certain 'normative appreciation' for public engagement (Marres 2012:54).

Significantly less attention has been paid, however, to exploring the particular political constitution of stakeholder engagement per se. This chapter addresses this lacuna by analyzing CANCERGEN's stakeholder engagement process through the lens of Thévenot's regimes of engagement framework,

with particular attention paid to the regime of *engaging in plans*. Here, I outline the characteristics of what I call a *pre-public platform*, which I argue operates according to a logic of containment that privileges local negotiations and tradeoffs among individual interests while working to preclude escalations to moments of public justification. Nevertheless, the chapter demonstrates a reflexive acknowledgment of a *horizon of publicity* -- that is, of some future possible moment where an issue can be made public -- on the part of the actors involved in these activities. This, I conclude, begs an approach to participation practices that considers the constant flow of situations between public moments and those that precede publicity, together constituting an *ecology of engagements and (e)valuations*.

4.4 Formalizing valuation in clinical research planning

While Chapter 4 focuses on the participation practices enacted in CANCERGEN, Chapter 5 addresses a still further element that was subsequently folded into into the wider prioritization dispositif: an economic-decision analytic modeling approach referred to as value of information (VOI) analysis. A primary use of VOI is to quantify, and sometimes even to monetize, the value to be gleaned from gathering further information on a given question as a means of ameliorating current decisional uncertainty. VOI analysis was an attractive tool for CANCERGEN project organizers because it promised a means of 'formalizing' the prioritization of diagnostic technologies and the design of clinical studies to be carried out within the Cooperative Group Program; it was viewed as a more durable quantitative adjunct to the open-ended, qualitative stakeholder deliberations that had been carried out in previous phases of the prioritization process. But what kind of 'value' does the single pecuniary output of a complex, multi-parameter VOI model signify within the CANCERGEN context?

To answer this question, I trace the ways in which project organizers drew from a number of conventional practices in the field of health economics to develop their models and adapted them in such a way that would maximize the likelihood of their usefulness in a locally circumscribed planning exercise. There were, in other words, a series of normative decisions made, on the one hand, about what the models would account for and how their results would be presented to the stakeholders; on the other hand, there was a simultaneous push to educate the stakeholders in such a way that they would be rendered adequate interpreters of the models' underlying logics.

In detailing this process, it becomes clear that despite the fact that the economized outputs of these models at first glance appear to be publicly qualified numbers corresponding to a *market worth* (Boltanski & Thévenot 2006; cf. Fourcade 2011a, 2011b), they are in fact indicative of some *other* mode of qualification. Looking to recent literatures in economic sociology and social studies of (e)valuation that take up themes of formalization and economization, I thus argue that the deployment of value of information analysis as a *formalizing practice* within CANCERGEN constitutes a mode of what I call *provisional economization* which works to *certify the worth* of competing avenues of public investment in carrying out clinical studies (cf. Çalışkan & Callon 2009; Lampland 2010; Thévenot 2015a). The functional character of these economic numbers and their use in streamlining the achievement of a stated objective -- of prioritizing and designing a trial for a diagnostic single technology -- points to a regime of *planning* rather than to any type of qualification according to the common good (e.g. Thévenot 2007). With reference to the *ecology of engagements and (e)valuations* described in the previous chapter, here too it is important to be cognizant of the ways in which these numbers may in fact translate into public concerns and modes of evaluation at some future moment. Having analytic tools such as those proposed in this chapter, which enable more holistic understandings of this flux, leaves analysts well positioned to account for the dynamics of larger scale shifts in policy over time (e.g. Muniesa et al. 2017).

4.5 Which technology to trial?

Chapter 6, the final empirical chapter of this thesis, draws on much of the same ethnographic material as the previous chapters and may thus occur to the reader as slightly empirically redundant. Analytically, however, the frame is in fact much wider than that of Chapters 4 and 5 as this penultimate chapter shifts to looking at the combinative deployment of stakeholder engagement and value of information analysis in the actual work of prioritizing and designing a comparative effectiveness study in CANCERGEN. The point here is thus unique: where the evaluative work in Chapter 4 relates to qualifications of *stakeholders* and explores how stakeholders themselves understood the value of their engagements, and in Chapter 5 pertains to the development of economic models in such a way that was envisioned as maximizing the uptake of model results by the stakeholders, this chapter homes in on how these valuation practices are themselves folded into the CANCERGEN prioritization process.

Here, we are returned to the questions with which this introductory chapter opened: Whom and what gets to count in the world? And how can we best use the conceptual tools of sociological inquiry to understand the practical as well as political implications bound up in asking and answering this question? Our answers to these questions in this chapter contribute to recent pragmatic and practice-based approaches to (e)valuation focused on questions about what is worth knowing, with particular focus on the prioritization and selection of a diagnostic technology to submit to the rigorous methodology of a randomized controlled comparative effectiveness research trial. Focusing on a central question at stake in CANCERGEN -- Which technology to trial? -- I position these qualitative-deliberative and quantitative-performative approaches as a series of *evaluative moments* within the project, each defined by specific evaluative equipments, agencies, and technologies. In tracing this syncretic approach to (e)valuation, I draw on Thévenot's regimes of engagement framework to contrast a *politics of elicitation* with a *politics of performation* which, despite both occurring within an ostensibly singular regime of planning, nevertheless signal the existence of a certain heterogeneity internal to the regime itself, evidenced by the quite distinct coordinative affordances and capacities for valorization that deliberative and quantitative modes of evaluation offer.

4.6 Conclusion

In this final chapter, I offer a summary and synthesis of the empirical and theoretical findings achieved in this thesis, focusing on the contributions they make to the wider sociological and STS literatures on participation, economization, and (e)valuation writ large. Additionally, I reflect on both some limitations of this research as well as possibilities for future work that it affords.

6. References

Abbott, A. (1988). *The System of Professions: An Essay on the Division of Expert Labor*. Chicago: University of Chicago Press.

Agre, P. E., & Chapman, D. (1990). What are plans for? *Robot. Auton. Syst.*, *6*(1–2), 17–34.

Auray, N. (2016). *L'Alerte ou l'enquête. Une sociologie pragmatique du numérique*. Paris: Presses des Mines/ParisTech.

Baszanger, I., & Dodier, N. (2004). Ethnography: relating the part to the whole. In D. Silverman (Ed.), Qualitative Research (2nd ed., pp. 9–34). London: SAGE.

Becker, H. (1986). *Doing Things Together: Selected Papers*. Evanston: Northwestern University Press.

Berg, M., & Timmermans, S. (2000). Orders and their others: on the constitution of universalities in medical work. *Configurations*, *8*(1), 31–61.

Bijker, W. E., Hughes, T. P., & Pinch, T. J. (Eds.). (1987). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT Press.

Birnbaum, H., & Slutsky, J. R. (2010). Guiding Comparative Effectiveness Research—A US Perspective. *Pharmacoeconomics*, *28*(10), 839–842.

Blokker, P., & Brighenti, A. (2011). An interview with Laurent Thévenot: on engagement, critique, commonality, and power. *European Journal of Social Theory*, *14*(3), 383–400.

Bohlin, I. (2012). Formalizing syntheses of medical knowledge: the rise of meta-analysis and systematic reviews. *Perspectives on Science*, *20*(3), 273–309.

Boltanski, L. (2011). *On Critique: A Sociology of Emancipation*. Cambridge, UK: Polity.

Boltanski, L., & Thévenot, L. (1999). The sociology of critical capacity. *European Journal of Social Theory*, *2*(3), 359–377.

------. (2006). *On Justification: Economies of Worth*. Princeton: Princeton University Press.

Bourret, P. (2005). BRCA Patients and Clinical Collectives: New Configurations of Action in Cancer Genetics Practices. *Social Studies of Science*, *35*(1), 41–68.

Bourret, P., Keating, P., & Cambrosio, A. (2011). Regulating diagnosis in post-genomic medicine: Re-aligning clinical judgment? *Social Science & Medicine*, *73*(6), 816–824.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT press.

Brook, R. H. (2009). Possible Outcomes of Comparative Effectiveness Research. *JAMA*, *302*(2), 194–195.

Califf, R. M. (2009). Clinical Research Sites—The Underappreciated Component of the Clinical Research System. *JAMA*, *302*(18), 2025–2027.

Çalışkan, K., & Callon, M. (2009). Economization, part 1: shifting attention from the economy towards processes of economization. *Economy and Society*, *38*(3), 369–398.

Callon, M., Lascoumes, P., & Barthe, Y. (2009). *Acting in an uncertain world: An essay on technical democracy*. Cambridge, MA: MIT Press.

Callon, Michel, Méadel, C., & Rabeharisoa, V. (2002). The economy of qualities. *Economy and Society*, *31*(2), 194–217.

Carpenter, D. (2010). Confidence Games: How Does Regulation Constitute Markets? In E. J. Balleisen & D. A. Moss (Eds.), *Government and Markets: Toward a New Theory of Regulation* (pp. 164–190). New York: Cambridge University Press.

Cefaï, D., Zimmermann, B., Nicolae, S., & Endreß, M. (2015). Introduction. *Human Studies*, *38*(1), 1–12.

Centemeri, L. (2010). The "conventional" objectivity of public space: how to think about the questionability of what we need to be unquestionable. Presented at the 2nd CES International Seminar on the Foundations of Economics Facts, Values and Objectivity, Centro de Estudos Sociais, Colégio de S. Jerónimo, Coimbra, PT.

------. (2015a). Reframing problems of incommensurability in environmental conflicts through pragmatic sociology: From value pluralism to the plurality of modes of engagement with the environment. *Environmental Values*, *24*(3), 299–320.

------. (2015b). Review of François Vatin (Ed.): Évaluer et valoriser: une sociologie économique de la mesure. *Human Studies*, *38*(1), 179–184.

Cheyns, E. (2011). Multi-stakeholder initiatives for sustainable agriculture: limits of the 'inclusiveness' paradigm. In S. Ponte, P. Gibbon, & J. Vestergaard (Eds.), *Governing through standards: Origins, drivers and limitations, eds. S. Ponte, S., P. Gibbon and J. Vestergaard* (pp. 210–235). Basingstoke: Palgrave Macmillan.

------. (2014). Making "minority voices" heard in transnational roundtables: the role of local NGOs in reintroducing justice and attachments. *Agriculture and Human Values*, *31*(3), 439–453.

Cochoy, F. (2014). From Strategy to Equipped Serendipity: Lessons from Ezio, the Black Angel of Florence. In R. M. Denny & P. L. Sunderland (Eds.), *Handbook of Anthropology in Business*. London: Routledge.

Dewey, J. (1939). Theory of valuation. *International Encyclopedia of Unified Science*, *II*(4), vii–66.

Dilts, D. (2010). US cancer trials may go the way of the Oldsmobile. *Nature Medicine*, *16*(6), 32.

Dodier, N., & Barbot, J. (2016). The Force of Dispositifs. *Annales. Histoire, Sciences Sociales: English Edition*, *71*(2), 291–317.

Dussauge, I., Helgesson, C.-F., & Lee, F. (Eds.). (2015). *Value Practices in the Life Sciences and Medicine*. Oxford: Oxford University Press.

Eliasoph, N., & Lichterman, P. (2003). Culture in interaction. *American Journal of Sociology*, *108*(4), 735–794.

Eranti, V. (2018). Engagements, grammars, and the public: From the liberal grammar to individual interests. *European Journal of Cultural and Political Sociology*, *5*(1–2), 42–65.

Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a Social Process. *Annual Review of Sociology*, *24*(1), 313–343.

------. (2008). A sociology of quantification. *European Journal of Sociology*, *49*(3), 401–436.

Evans, B. J., Burke, W., & Jarvik, G. P. (2015). The FDA and Genomic Tests — Getting Regulation Right. *New England Journal of Medicine*, *372*(23), 2258–2264.

FCC-CER [Federal Coordinating Council for Comparative Effectiveness Research]. (2009). *Report to the President and the Congress, June 30, 2009*. Washington, DC: The Council.

Fourcade, M. (2011a). Cents and sensibility: Economic valuation and the nature of "nature" 1. *American Journal of Sociology*, *116*(6), 1721–1777.

------. (2011b). Price and Prejudice: On Economics, and the Enchantment/Disenchantment of Nature. In J. Beckert & P. Aspers (Eds.), *The Worth of Goods. Valuation and Pricing in the Economy*. Oxford, England: Oxford University Press.

Garber, A. M., & Tunis, S. R. (2009). Does Comparative-Effectiveness Research Threaten Personalized Medicine? *New England Journal of Medicine*, *360*(19), 1925–1927.

Gieryn, T. F. (1999). *Cultural Boundaries of Science: Credibility on the Line*. Chicago: University of Chicago Press.

Hansen, M. P. (2016). Non-normative critique: Foucault and pragmatic sociology as tactical re-politicization. *European Journal of Social Theory*, *19*(1), 127–145.

Hedgecoe, A. (2004). *The Politics of Personalised Medicine: Pharmacogenetics in the Clinic*. Cambridge, UK: Cambridge University Press.

------. (2008). From Resistance to Usefulness: Sociology and the Clinical Use of Genetic Tests. *BioSocieties*, *3*(2), 183–194.

Helgesson, C.-F. (2016). Folded Valuations? *Valuation Studies*, *4*(2), 93–102.

Hoffman, A. S., Cambrosio, A., & Battista, R. (2016). Comparative Effectiveness Research in Health Technology Assessment. In A. Levy & B. Sobolev (Eds.), *Comparative Effectiveness Research in Health Services* (pp. 57–93). New York: Springer.

Hogarth, S. (2015). Neoliberal technocracy: Explaining how and why the US Food and Drug Administration has championed pharmacogenomics. *Social Science & Medicine*, *131*, 255–262.

Hughes, C. E. (1984). *The Sociological Eye: Selected Papers*. New Brunswick: Transaction Publishers.

IOM [Institute of Medicine]. (2009). *Initial National Priorities for Comparative Effectiveness Research*. Washington, DC. Retrieved from http://nationalacademies.org/hmd/Reports/2009/ComparativeEffectivenessResearchPriorities.aspx

Keating, P., & Cambrosio, A. (2012). *Cancer on Trial: Oncology as a New Style of Practice*. Chicago: The University of Chicago Press.

Kessous, E. (2015). The Attention Economy Between Market Capturing and Commitment in the Polity. *Œconomia. History, Methodology, Philosophy*, (5–1), 77–101.

Khoury, M. J. (2010). Dealing with the evidence dilemma in genomics and personalized medicine. *Clinical Pharmacology & Therapeutics*, *87*(6), 635–638.

Knaapen, L., Cazeneuve, H., Cambrosio, A., Castel, P., & Fervers, B. (2010). Pragmatic evidence and textual arrangements: A case study of French clinical cancer guidelines. *Social Science & Medicine*, *71*(4), 685–692.

Kohli-Laven, N., Bourret, P., Keating, P., & Cambrosio, A. (2011). Cancer clinical trials in the era of genomic signatures: biomedical innovation, clinical utility, and regulatory-scientific hybrids. *Social Studies of Science*, 0306312711398741.

Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, *38*, 201–221.

Lampland, M. (2010). False numbers as formalizing practices. *Social Studies of Science*, *40*(3), 377–404.

Levy, A. R., & Garrison, L. P. (2010). Apples and Oranges? Assessing Comparative Effectiveness and Comparative Value in the US and Other Countries. *Value in Health*, *13*(s1), S1–S3.

Luce, B. R., Drummond, M., Jönsson, B., Neumann, P. J., Schwartz, J. S., Siebert, U. W. E., & Sullivan, S. D. (2010). EBM, HTA, and CER: clearing the confusion. *Milbank Quarterly*, *88*(2), 256–276.

Marks, H. M. (2009, unpublished). What does the FDA do? Regulation, Drug Markets and Medical Practice, 1906-2009.

Marres, N. (2012). *Material Participation: Technology, the Environment and Everyday Publics*. Basingstoke: Palgrave Macmillan.

Moreira, T. (2005). Diversity in clinical guidelines: the role of repertoires of evaluation. *Social Science & Medicine*, *60*(9), 1975–1985.

------. (2012). Health Care Standards and the Politics of Singularities Shifting In and Out of Context. *Science, Technology & Human Values*, *37*(4), 307–331.

Muniesa, F. (2012). A flank movement in the understanding of valuation. *The Sociological Review*, *59*(s2), 24–38.

Muniesa, F., Doganova, L., Ortiz, H., Pina-Stranger, Á., Paterson, F., Bourgoin, A., … Yon, G. (2017). *Capitalization: A Cultural Guide*. Paris: Presses des Mines.

NCI [National Cancer Institute]. (2016). An Overview of NCI's National Clinical Trials Network - National Cancer Institute. Retrieved March 30, 2016, from https://www.cancer.gov/research/areas/clinical-trials/nctn/#history

Normanno, N., Tejpar, S., Morgillo, F., De Luca, A., Van Cutsem, E., & Ciardiello, F. (2009). Implications for KRAS status and EGFR-targeted therapies in metastatic CRC. *Nature Reviews Clinical Oncology*, *6*(9), 519–527.

Nowotny, H., Scott, P. B., & Gibbons, M. T. (2001). *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge, UK: Polity Press.

Quéré, L. (1998). The still-neglected situation? *Reseaux. The French Journal of Communication*, *6*(2), 223–253.

------. (2015). Value as a Social Fact: An Adverbial Approach. *Human Studies*, *38*(1), 157–177.

Scoggins, J. F., & Ramsey, S. D. (2010). A national cancer clinical trials system for the 21st century: reinvigorating the NCI Cooperative Group Program. *Journal of the National Cancer Institute*, *102*(17), 1371–1371.

Silva-Castañeda, L. (2012). A forest of evidence: third-party certification and multiple forms of proof—a case study of oil palm plantations in Indonesia. *Agriculture and Human Values*, *29*(3), 361–370.

------. In the shadow of benchmarks. Normative and ontological issues in the governance of land. *Environment and Planning A: Economy and Space*, *48*(4), 681–698.

Sox, H. C. (June). Defining Comparative Effectiveness Research: The Importance of Getting It Right. *Medical Care*, *48*(6), S7.

Stark, D. (2009). *The sense of dissonance: Accounts of worth in economic life*. Princeton: Princeton University Press.

Sturdy, S. (2018, under review). Seeing utility: regulatory reform and genetic tests in the USA, 1989-2000. *Social Science & Medicine*.

Tavory, I., & Eliasoph, N. (2013). Coordinating Futures: Toward a Theory of Anticipation. *American Journal of Sociology*, *118*(4), 908–942.

Thévenot, L. (1984). Rules and implements: investment in forms. *Social Science Information*, *23*(1), 1–45.

------. (1990). L'action qui convient. In P. Pharo & L. Quéré (Eds.), *Les formes de l'action. Sémantique et sociologie.* (pp. 39–69). Paris: Éditions de l'EHESS.

------. (1995). L'action en plan. *Sociologie Du Travail*, *37*(3), 411–434.

------. (2006). Institutions and agency: Differentiating regimes of engagement. Presented at the First Max Planck Summer Conference on Economy and Society.

------. (2007). The plurality of cognitive formats and engagements moving between the familiar and the public. *European Journal of Social Theory*, *10*(3), 409–423.

------. (2014a). Engaging with the politics of participative art in practice. In T. Zembylas (Ed.), *Artistic Practices: Social Interactions and Cultural Dynamics, 1st Edition* (pp. 132–150). London: Routledge.

------. (2014b). Enlarging Conceptions of Testing Moments and Critical Theory: Economies of Worth, On Critique, and Sociology of Engagements. In S. Susen & B. S. Turner (Eds.), *The Spirit of Luc Boltanski: Essays on the "Pragmatic Sociology of Critique"* (pp. 245–261). London: Anthem Press.

------. (2015a). Certifying the World: Power Infrastructures and Practices in Economies of Conventional Forms. In P. Aspers & N. Dodd (Eds.), *Re-Imagining Economic Sociology* (pp. 195–216). Oxford: Oxford University Press.

------. (2015b). Making Commonality in the Plural on the Basis of Binding Engagements. In M. Dumouchel & R. Gotoh (Eds.), *Social Bonds as Freedom: Revisiting the Dichotomy of the Universal and the Particular*. Berghahn Books.

Timmermans, S., & Berg, M. (2003). *The gold standard: The challenge of evidence-based medicine and standardization in health care*. Philadelphia: Temple University Press.

Timmermans, S., & Epstein, S. (2010). A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization. *Annual Review of Sociology*, *36*(1), 69–89.

Twombly, R. (2009). Researchers Left To Guess at Outcomes of Most Cancer Clinical Trials. *JNCI: Journal of the National Cancer Institute*, *101*(2), 72–74.

Vatin, F. (2013). Valuation as Evaluating and Valorizing. *Valuation Studies*, *1*(1), 31–50.

Weisz, G., Cambrosio, A., Keating, P., Knaapen, L., Schlich, T., & Tournay, V. J. (2007). The emergence of clinical practice guidelines. *Milbank Quarterly*, *85*(4), 691–727.

Westenholz, A. (2012). *The Janus face of commercial open source software communities: An investigation into institutional (non)work by interacting institutional actors*. Copenhagen: Copenhagen Business School Press.

Chapter 2

Materials & methods

1. Combining sites, narrowing the scope

This thesis is composed of a qualitative analysis of primary and secondary empirical data collected over the span of approximately six years (2009-2015). During this period, I employed three distinct data collection methodologies: ethnographic (participant)-observations at conferences, workshops, and meetings (n=13); in-depth, semi-structured interviews with key informants (n=27); and close documentary analysis of primary and secondary texts. As I describe below, the precise scope of the project shifted considerably over time, which can be roughly grouped into three phases of research, each with its own ethnographic posture and a corresponding set of data collection and analysis techniques.

2. Surveying the CER landscape

During the initial phase of research, my focus was quite broad as I was surveying the emerging nexus of two novel fields: comparative effectiveness research and personalized medicine. The guiding question here was how comparative effectiveness research was being situated, both historically and definitionally, as well as how it was coming to structure conversations about the regulation and evaluation of personalized medicine technologies. Two early interviews, including the aforementioned interview with Dr. Forrester, as well as a second with another clinician-researcher in New York City who had been funded to carry out a project on comparative effectiveness research and personalized medicine, provided some early leads for further contextualization of the field.

Universities, government agencies, and nonprofit organizations were also hosting meetings during this time, sometimes addressing one or the other of these two domains and frequently overlaps among them. Thus, ethnographic observation was a key tool for collecting data to understand various actors' rhetorical framings of CER and personalized medicine, their articulations of what such a research program could and should look like, and for unpacking how various issues at the intersection of CER and personalized medicine were being problematized by actors in this space.

I conducted ethnographic fieldwork at these earlier CER-focused events (n=2), the first of which was a daylong cross-disciplinary workshop, held in May 2010 at the Radcliffe Institute in Cambridge, MA. This was a gathering of an international group of historians, social scientists, ethicists, policy experts, clinicians, and researchers who met to discuss the goals, methods, and practices of comparative effectiveness research. The meeting was a key site for sensitizing me to the ways in which these actors were respectively envisioning the past, present, and future of this emerging mode of research. The opening remarks at the event were especially revealing: over a year after the Stimulus Package was signed into law which apportioned over US$1 billion for CER, there was still an understanding that a set of 'basic questions' remained 'fundamentally unanswered': What is CER? How does one define it? How does one going about doing it? And how should the results of CER studies be applied? One attendee noted: 'There is a globalized form of evaluation progressing from basic [research] to RCT [randomized controlled trials] to post-market trials, now diffused around the world,' while by contrast, comparative effectiveness 'has no standardized template -- although many are in the works.'

Several months later, at the second of these events, I found myself sitting in a large auditorium on the main campus of the National Institutes of Health in Bethesda, MD. It occurred to me that these questions had not yet been established as 'matters of fact' but were still unsettled 'matters of concern' (Latour 1987). In fact, in light of the theme of the conference -- which was to address the 'essential interface' of comparative effectiveness research and personalized medicine -- there were many more nuances than simply defining the contours of what a CER research program looks like. As NIH Director Francis Collins said during his keynote address on the first day of the meeting: 'I've been in many situations and had the question posed, about comparative effectiveness research and personalized medicine -- because there are certainly some of those who pose the question, who think there might be a conflict between these two. I don't share that.' As it turned out, nearly all of the presentations over the course of the two-day meeting were geared precisely towards demonstrating the *compatibility* of these approaches, and drew from a number of historical examples of clinical research programs that have demonstrated this complementarity.

One particular presentation by Robert Croyle, Director of the National Cancer Institute's Division of Cancer Control and Population Sciences, stood out to me as particularly relevant to my

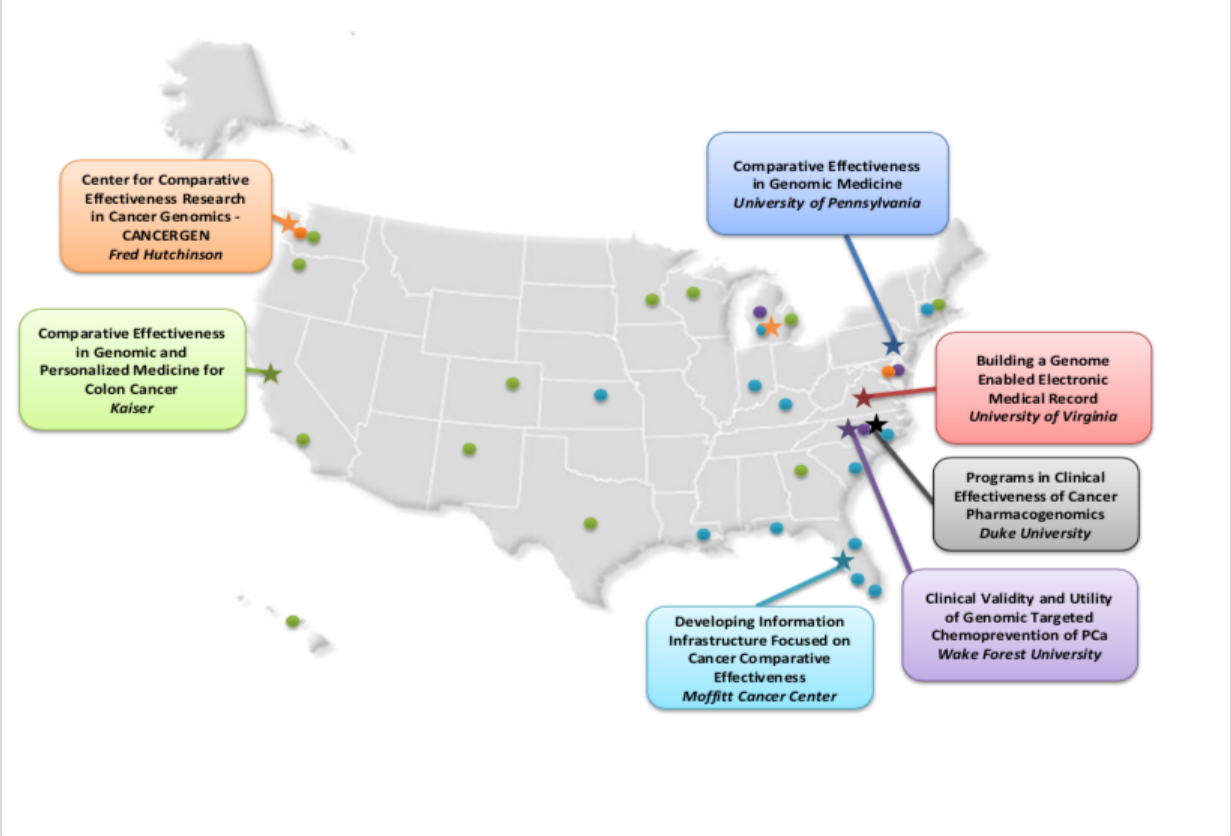
interests in evidentiary practices in oncology. His talk outlined the ongoing 'revamping' of the national Cooperative Group clinical trials system, pointing out that the reforms that were then underway sought to improve operational efficiency and prioritization, increase funding for complex trials, as well as to 'increase [the] role of non-NCI scientists and stakeholders in governance' of these processes. As he went on to say:

> all of these aspects link also to comparative effectiveness research because the issue discussed earlier -- being one of cost and management. There's no way we can do all the trials we want to do, especially the state of the art trials that incorporate biospecimen collection, assays, genomics, etcetera. And therefore in the era of personalized medicine it's even more important that we have criteria and processes for prioritizing trials. [...] If we are going to do these more sophisticated trials and figure out how to pay for them, we need proof of principle projects and studies, and there are some of those that are underway.

Croyle's main point posed an interesting inversion of what appeared to be the more typical CER discourse of the time, which stressed the need for medical *evidence* to inform clinical *practice*. What is also needed, he argued, is '*practice* informing *evidence development* and the development of prioritization of research questions at funding entities like the NIH.' The discourse deployed here thus shifts the focus upstream from concerns about designing evidentiary tools -- in this case, randomized controlled trials -- to a more 'meta' discourse (cf. Cambrosio et al. 2006) centered on developing evidence that can inform the design of those tools, e.g. by first designing new prioritization criteria and processes.

Advancing to a slide with a map of the United States (Figure 2.1), Croyle went on to provide a high level overview of a $50M investment in a consortium of seven projects that 'marry comparative effectiveness research in the area of genomics and personalized medicine,' and which he described as 'trying to kind of link together latest best evidence in terms of molecular oncology and address issues of implementation, uptake, dissemination, and *evaluation* in practical healthcare settings.' These were the aforementioned 'proof of principle' and 'capacity development' projects, whose support was provided by a new line of funding called Grand Opportunities (henceforth GO Grants), a grant mechanism that the NIH had developed for alloting a portion of its US$400m of ARRA CER funding.

Figure 2.1 Map of NCI GO Grant Consortium in CER and Personalized Medicine



(Presented by Robert Croyle at NCI, October 2010)

3. Multi-sited ethnography in a national CER consortium

With Croyle's presentation came the passage into a second phase of research, where I decided to narrow the focus of my fieldwork to these seven GO Grants. Here I sought to understand how specific evaluative practices were being used to *enact* (e.g. Mol 2002) CER in the domain of genomics and personalized medicine in cancer care as well as to understand what it meant at these specific sites for practice to inform the development of evidence and prioritization. Rather than surveying the discursive landscape of CER as I had done during the first phase, this second phase took the shape of a *multi-sited ethnography* approach, which 'centers attention on the construction of the ethnographic object' (Hine 2007:655). A key element of this approach is the importance placed on

the ethnographer not equating distinct field sites to 'culturally distinct wholes' but instead 'mak[ing] it part of their goal to find out where interesting things might be going on' (ibid.:661). Indeed, there were *many* interesting things going on within the seven GO Grant consortium projects to which I had turned my attention (Simonds et al. 2013).

For instance, two projects were working to develop and build out technical capacities of information infrastructures by embedding patients' genomic information and other relevant data into electronic health record systems, which could further facilitate comparative effectiveness research by helping to categorize patient populations for comparison in a clinical study of genomic-guided interventions (e.g. Fenstermacher et al. 2011). Another subset of projects were more focused on clinical research infrastructures, including developing novel ways of measuring and assessing patient outcomes in personalized medicine and constructing repositories for biological specimens that could be used to comparatively analyze patient responses to drug treatments (e.g. Ginsburg & Kuderer 2012). As for CANCERGEN, it incorporated a number of infrastructural and clinical elements, but also seemed to stand out from the rest for its emphasis streamlining the prioritization and design of comparative effectiveness studies, and whose engagement of stakeholders in this process stood out as a defining feature of CER.

During this second phase of my research, I conducted a total of 25 semi-structured, in-depth interviews with individuals participating in in the GO Grant consortium projects, as well as in other CER-related activities. Interviews with CANCERGEN project organizers and members of the CANCERGEN External Stakeholder Advisory Group comprised nearly half of these interviews (n=10). The former group included the project's Principal Investigator and a Project Director, both based at the academic cancer research center (n=1); the CANCERGEN stakeholder engagement lead (n=1), based at the nonprofit organization specializing in stakeholder engagement (n=1); project organizers from the public university who co-led the economic modeling component (n=2, one of whom was a project co-PI); and a statistician (n=1, also a co-PI) from SWOG, the clinical trial cooperative group who was active in designing the prospective CER studies. Counted among the ESAG members I interviewed (n=4) were a representative from a large national advocacy umbrella organization focused on accelerating health benefits accrued from genomic research, representing the 'consumer' perspective (n=1); the CEO of a diagnostics manufacturing firm, invited

to represent the 'industry' perspective; an individual employed by a private regional health insurance company (n=1), representing the 'private payer' perspective; and the director of a publicly-funded state health technology assessment agency (n=1), representing the 'public payer/policy' perspective.

I was able to conduct fieldwork, including an additional set of interviews (n=12) and (participant-)observations at meetings (n=3) associated with three of the other clinically-oriented GO Consortium projects. This included interviews with principal investigators (n=2) and attendance at two project team meetings (n=2) at a university hospital-based project carrying out pilot projects in evidence generation and synthesis in the areas of breast and lung cancer; interviews with principal investigators (n=3) and attendance at a team meeting (n=1) at a second university hospital-based project which sought to assess the clinical effectiveness of pharmacogenomics in oncology; and interviews with several participants (n=7), including principal investigators (n=2) and associated project and research staff (n=5), based at the research arm of a regional private health maintenance organization who were conducting analyses on the effectiveness of personalized medicine technologies in the detection and treatment of colorectal cancers.

Over this same period of time I attended Consortium-wide meetings (n=2) as a (participant-)observer where I was able to observe discussions among team members relevant to the scoping, organization, prioritization, progress, and future visions of their ongoing work. The first was a two-day GO Consortium workshop, whose first day was focused on gauging stakeholders' perceptions of biomedical evidence in CER and personalized medicine (cf. Deverka et al. 2012c) and the second day consisted of an all-hands meeting for project PIs and staff to give updates on their progress and to work collectively on drafts for group publications. The second of these meetings was held at the NIH towards the end of the funding period for the GO consortium, where the objective was to once again share progress of the projects with the Consortium community and program officers and to plan possible next steps and future funding opportunities. Together, these lent me a wider perspective on what was unfolding within the space of the Consortium writ large.

I attended an additional set of meetings and workshops (n=4) as well during this second phase, although these were somewhat broader in scope: a summit on CER in oncology (n=1), hosted by the same nonprofit healthcare stakeholder engagement organization that participated in CANCERGEN; meetings of the International Society for Pharmaceutical and Outcomes Research

(n=2), where GO Grant project participants presented their ongoing research findings; and a public Board of Directors meeting (n=1) of the Patient Centered Outcomes Research Institute (PCORI), the non-governmental organization whose founding was mandated as a provision of President Obama's 2010 Patient Protection and Affordable Care Act and which was tasked with supporting comparative effectiveness research at the national scale. I also conducted additional interviews (n=3) with members of the then newly-formed PCORI methodology committee (n=2), and with a principal investigator from a non-genomics Grand Opportunities grant in oncology (n=1), once again with the objective of widening my understanding of how the GO Consortium fit within the wider ecology of CER activities which were being rolled out at this time.

4. Combinative ethnography in CANCERGEN

The final phase of my research underwent a reorientation: what was a *multi-sited* approach in phase two augmented during this third phase as a *combinative ethnography* (Baszanger & Dodier 2004). Akin to the constant comparison approach that originated with the Chicago School of sociology and which continues to inform much of the grounded theory tradition of sociological analysis (Clarke 2005; Glaser & Strauss 1973; Star 2010), the *sociological pragmatics* of combinative ethnography leads away from totalization and towards the accretion of single cases, analyzed 'as a combination between different logics of action that coexist not only in the field under consideration, but even within these individuals or during their encounters' (Baszanger & Dodier 2004:19).

That the transition entailed a *narrowing* of empirical focus to only a single GO Grant may at first glance seem paradoxical, but produced certain affordances for my analysis, namely the opportunity to try and reconcile the paradox by exploring the potentialities of studying an ostensibly singular project using combinative methods. My decision to focus only on a single project was, however, ultimately motivated by practical considerations. The wider GO Grant Consortium was a complex program whose seven constituent units were in fact each made up of multiple sub-projects, themselves addressing a myriad of (sometimes competing) objectives and themes in their own right. Being empirically and analytically faithful to the totality of the work I was observing, even in one small corner of the wider field of CER, thus occurred to me as an exercise in frustration if not outright futility. This is hardly surprising, as finding a site for studying a particular phenomenon or

type of action often reveals the field 'to be more disparate than anticipated' (Baszanger & Dodier 2004:20).

Shared by both multi-sited and combinative ethnographic approaches is a common appreciation for openness and diversity. Hine emphasizes a 'focus on diversity as the key insight' of multi-sited ethnography and calls on researchers to 'remain more ambivalent about relevant locations' in the field in such a way they remain open to noticing heterogeneity (2007:668; cf. Tsing 2015). Baszanger and Dodier similarly proclaim the necessity for openness such that the combinative ethnographer may 'discover the elements making up the markers and the tools that people mobilize in their interactions with others and, more generally, with the world' (2004:11). There is, however, also an important distinction between the two approaches. Multi-sited ethnography often entails traveling 'to many different places to explore different aspects of [a] phenomenon' (Hine 2007:666), the analysis of which focuses on 'tracing and describing the connections and *relationships among sites* thought previously incommensurate' (Marcus 1998 quoted in ibid.:656, italics mine). Combinative ethnography, meanwhile, sets the researcher on the path of developing 'a combinative inventory of possible *situations*' as the researcher 'circulates between several sites, depending on which dimensions appear relevant in the analysis of each case' (Baszanger & Dodier 2004:19, italics mine). Analysis proceeds by 'demonstrating the means by which actors navigate between, and oftentimes combine, multiple types of activity, commitments, and resources as they travel through [these] various *situations*,' and becomes 'gradually enriched by new examples displaying new forms of activity and patterns of articulation' (ibid., italics mine).

In other words, while the level of analysis in multi-sited ethnography is trained on the multiple sites themselves and the work of making them commensurate vis-a-vis a given phenomenon, the combinative approach is far more attuned to accounting for the cascade of *situations* (cf. Goffman 1964; Quéré 1998) in which actors find themselves and the means by which they negotiate transitioning through these instances. Given the description provided above in Chapter 1, Section 2.4, the extent of organizational, temporal, and material variegation internal to CANCERGEN alone (to say nothing of the other six Consortium projects) -- in terms of activities, commitments, and resources with which the field presented me -- should not be lost on the reader, and certainly was not lost on me as an ethnographer.

For one, the project was organizationally complex, involving a number of institutions as well as the External Stakeholder Advisory Group. Members of the ESAG were invited to participate in the project specifically because of their differently located positionalities within the healthcare field and were thus perceived as possessing a set of competing commitments; rather than viewed as a stumbling block, however, these this diversity of individual perspectives and opinions was viewed as an asset to the project. As well, the project had a specific temporal horizon, a three-year funding period during which the prioritization and design process unfolded in a series of steps, or *evaluative moments* (see Chapter 1, Section 4.5 and Chapter 6; cf. Ehrenstein & Neyland 2016). At each step along the way, particular combinations of actors, objects, and evaluative techniques were deployed, largely in relation to the linked approaches of stakeholder engagement and value of information analysis, which were themselves enacted in different ways during these various moments. In line with the precepts of the combinative approach, my data consisted of an *ethnographic casebook* describing how actors dealt with the harmonious and/or conflictual situations that arose from their *pragmatic flexibility* as they moved through these various evaluative moments.

Proceeding in this way contributed to building out a 'sociology of encounters' between project participants over the course of the project and to understand the common 'fund of skills offering various possibilities of commitment in the encounter[s], while remaining within the constraints fixed by the arrangements created by the initial situational context' (Baszanger & Dodier 2004:24). This was, in turn, informed by my casebook whose contents drew from an extensive engagement with data collected during the aforementioned (participant-)observations and interviews, as well as from analyzing textual artefacts, including digital traces (Geiger & Ribes 2011). Manuscripts and abstracts, authored by project participants and published in academic journals and conference proceedings, were an important data source here, as were other inscriptions and scalar devices (Latour 1987; Ribes 2014), produced and circulated internally amongst CANCERGEN project organizers and stakeholders: workshop agendas, summaries, and informational materials; powerpoint slides; evaluative metrics; and survey questions and results. A third source of textual artefacts were transcripts of the two in-person External Stakeholder Advisory Group meetings that the project hosted over the course of its three-year funding period, one of which I attended in person and the other which had occurred very early on in the project.

Triangulation of these three data sources aided in reconstructing or otherwise elaborating upon those most analytically interesting situations, as well as in articulating actors' grammars -- that is, their ways of *talking* about and justifying their positions and actions, whether as future visioneering or post-hoc accounts -- together with their actual *practices* (e.g. Eliasoph et al. 2018). In this regard, triangulation also works as a strategy of trustworthiness and as a means of checking the veracity and consistency of fieldwork data, allowing the analyst to see how information is (re)presented across different spaces and venues. For example, the analyst may ask how a post-hoc account of a situation (like a project meeting) given by a respondent during an interview measures up with how the situation was recorded on a transcript of that meeting, how it was framed in a published article, or otherwise summarized in meeting minutes. If one finds inconsistencies, it is then possible to probe further to understand why this was the case, which can be both empirically and conceptually revealing.

In the case of CANCERGEN, the very existence certain textual artefacts such as meeting transcripts points to a curious feature of the project: in addition to functioning as a *platform* for prioritizing and designing comparative effectiveness research studies, it was also something of an *experiment* in the deployment and integration of stakeholder engagement and value of information analysis (e.g. Carlson et al. 2018; cf. Lezaun et al. 2017; Muniesa & Callon 2007). As these activities unfolded, project organizers were studying them as research objects in their own right. Several of CANCERGEN's published outputs were process-based descriptions of the design and deployment of these different methods, based on qualitative analyses of meeting transcripts and interviews that project organizers themselves conducted with stakeholders at different points throughout the project (Carlson et al. 2013; cf. Deverka et al. 2012b).

But despite what the actors themselves may analyze and publish in their own accounts of these evaluative situations, combinative ethnography seeks to render publicly available 'the elements constituting the (often hidden) pragmatic condition of individuals' for which durable traces are rarely left (Baszanger & Dodier 2004:27; cf. Latour & Woolgar 1986). This version of *uncovering* can be counterposed to engaging in explicit *critique*, where the analyst is assumed to possess some additional capacity for reflexivity over and beyond the actors whom they are observing. What typically results

from this latter posture are analyses focused on a *critical unveiling* of whatever 'hidden reality' the actors are not themselves somehow conscious of or capable of grasping (Latour 2005:32-33).

Assuming a more pragmatic ethos, the analyst is equipped to take seriously the practical activities and moral-cognitive framings they encounter in the field, and to notice the means by which actors organize and the resources upon which they draw when critiquing one another (Hansen 2016). This proves especially important when studying (e)valuation practices since it is precisely the actors' own evaluative techniques and judgments that are the objects of analysis. As well, as we saw in Section 3 of Chapter 1, there is in fact a central tendency towards the latter stance within the growing corpus of contemporary studies of (e)valuation (Dussauge et al. 2015; Antal et al. 2015; Kornberger et al. 2015), what one may even describe as a profusion of *pragmatist* or otherwise *pragmatic* approaches to valuation (Asdal 2018; Holden et al. 2013; Stark 2009; Stavo-Debauge 2012).

5. A note on conceptual innovation in the shadow of grounded theory

Before moving on with the remainder of the thesis, it begs noting that much of its originality (Guetzkow et al. 2004) in terms of the central theoretical and conceptual contributions I propose in the following chapters -- in particular, the notions of *pre-public platforms*, *provisional economization*, *certified worth*, and the *ecology of engagements and (e)valuations* -- are the outcomes of a process inspired by the aforementioned grounded theory approach. Herein, I offer a brief narrative about how these concepts were developed over the course of my research such that the reader may better appreciate how something that began as a (mostly) opportunistic entry into a new field site transitioned into a meditation on how we might understand the contours and limits of public as well as other-than-public modes of engagement and their corresponding politics of coordination and evaluation.

Although it was apparent to me in my initial encounters with CANCERGEN that stakeholder engagement would be an important element that project, I had no intention of building any kind of theory about such an arrangement nor of its link to wider questions of publicity. And yet in the earliest phases of analyzing my data, I encountered the distinction which CANCERGEN project organizers drew between a category they labeled 'stakeholders' on the one hand, and the categories of 'public' and 'citizen' on the other. In some of my earliest writing about the project, including in

Chapter 3 of this thesis, I had an intuition that holding these categories apart was somehow important.

For instance, I noted the 'conceptual slippage' between the use of notions of 'the public' and use of the the term 'stakeholder' one sometimes finds in the extant literature on participation practices in healthcare and argued that while the terms aren't necessarily mutually exclusive, 'neither should they be used interchangeably' (Hoffman et al. 2016:70). In making such a claim, I drew from my interlocutors' own terms, who distinguished the former from the latter insofar as the latter's self-interest in a given issue and their ability to render future actions both more legitimate and of a higher quality as opposed to the ostensibly disinterested (or even uninterested) members of the former category (Deverka et al. 2012b, cf. Lehoux et al. 2012). Confronted with this distinction, I was led to further probe what difference it made for the actors themselves in relation to how they went about coordinating the work carried out over the course of the project.

In many ways, my process of coming to grips with what a stakeholder was within the setting of CANCERGEN -- an early step in developing a wider theory of an *ecology of engagements and (e)valuations* -- closely mirrors the account that interactionist sociologist Howard S. Becker (1993) gives in his amusing treatise on grounded theory, 'How I learned what a crock was.' Here, he details his own efforts to understand the meaning of a 'crock' -- a term used by the group of medical trainees he was studying in the mid-1950s to describe patients who presented with psychosomatic illness. Over the course of testing the meaning of the term by using it in different ways and gauging the responses of his medical resident interlocutors, he concluded that there were three main things that the trainees most valued in their work on the hospital ward: the ability to learn 'the sights, sounds and smells of disease in a living person,' or what he labels 'clinical experience'; to learn that clinical experience across a diversity of maladies in the most expeditious way possible; and to have the opportunity to perform 'medical miracles,' or what he terms the 'medical responsibility' perspective (ibid:33-4).

Crocks presented no such opportunities for the trainees and thus garnered the their annoyance, and an accompanying label through which they expressed it. As Becker contends in his paper, '[w]hen members of one status category make [...] distinctions among the members of another status category with whom they regularly deal, the distinction will reflect the interests of the members of the first category in the relationship' (ibid.:31). More to the point, he adds, those distinctions which

members of the former group make between other categories of actors also reflect the interests they are seeking to 'maximize' and thus what they hope to get out of those relationships (ibid.) -- that is, the kind of 'evaluative good' they are pursuing (cf. Thévenot 2007). Important for the present discussion is the fact that Becker's (1993) grounded theory of a 'crock' was not a 'lightning bolt of intuition' but was rather 'guided by sociological theorizing every step of the way' (31).

In my own fieldwork, what CANCERGEN project organizers sought in their work of conducting a stakeholder engagement exercise (rather than involving the 'public' or 'citizens') was a rational and efficient approach to prioritizing and designing CER studies of personalized medicine technologies in oncology, as indicated in their very operationalization of the term. Involving stakeholders, understood as interested actors able to contribute to a more rational prioritization and design process, were thus valuable to project organizers because they were the very actors upon whom the project's success depended. This would also enable project organizers to demonstrate that their novel process was an improvement over the more 'science-based' approach to prioritizing clinical studies, while the stakeholders' inputs into the prioritization and design of studies would also mean that the outputs of those studies would be of a higher quality and more useful to a greater number of actors within the healthcare system.

While thinking through what this all amounted to, and drawing on my continued engagement -- with my fieldwork and interviews, with analyzing the various documents produced by CANCERGEN project participants, as well as my ongoing surveying of the sociological scholarship on engagement and evaluation -- I began to think that there was a more general point to be made about the texture of CANCERGEN, its coordinative infrastructure, and the coincident modes of (e)valuation I was noticing vis-a-vis distinguishing publicly qualified phenomena from what might be other-than-public. It was around this time that I stumbled upon Thévenot's *regimes of engagement* (ROE) framework for the first time. Initially, my interest was piqued on account of his own attempt at grappling with the figure of the 'stakeholder,' which he posits as the prototypical mode of agency within what he characterizes as the regime of engaging in plans (previously discussed in Chapter 1, Section 3.4).

The parallels between how CANCERGEN project organizers and Thévenot himself characterized the figure of the 'stakeholder' struck me as uncanny, as did how each differentiated the

stakeholder from some other publicly qualified category, such as 'citizens' or members 'the public.' This led me to further probe the ROE framework to see if there were other elements that resonated with my fieldwork data, especially in terms of the planning regime. Indeed, the deeper I looked into the theory, the more resonance and synergy I found with the empirical bits I had been collecting about CANCERGEN's participation practices -- it was not just the particular formatting of stakeholder agency that corresponded here, but also the broader evaluative good of accomplishing a stated goal of action (choosing a single trial, and then developing a design for that study); the types of functional information that was developed and deployed over the course of the project; and the actors own reflexive acknowledgment that what they were doing was of a more confined nature than engaging in public critiques, while also recognizing various ways of publicizing the outcomes of their work at some future moment. Together, these features contributed to the development of a concept I call *pre-public platforms*, of which I argue CANCERGEN stands as a paradigmatic example and is the central focus of Chapter 4 of this thesis.

From there, I then began to look back across the rest of my data, beyond participation practices alone, to see if there were other cases pointing to pre-public forms of coordination and evaluation -- a kind of data analysis phase version of what Glaser and Strauss (1973) call 'theoretical sampling.' Just as I had been preoccupied by the peculiar form of participation I was observing with CANCERGEN, so too did another significant focus of the project leave me somewhat unsettled: the aforementioned use of value of information analysis (VOI). The very process of assigning an economic value to a particular trial testing the usefulness of a particular personalized medicine technology was foreign to me at that stage. Moreover, I was in fact a bit surprised to encounter this approach within what I understood to be a participation exercise in the first place. But given its central role in CANCERGEN's wider process, I began to look into what VOI models were all about, how they were constructed, what phenomena they accounted for, and how they were deployed in the prioritization and design process. The closer I got to the models and their economized outputs, the more unfamiliar did their pecuniary values seem to me.

It was only by scrutinizing the various calculative conventions that my interlocutors described to me during my interviews with them, and which they discussed in their publications about VOI modeling, that I began to notice their correspondence with CANCERGEN's wider coordinative

infrastructure and certain parallels to how stakeholder participation was qualified in that setting. For instance, the models incorporated various types of economic phenomena that under other circumstances often take on a public quality and are publicly legible -- like the use of cost-effectiveness thresholds and willingness-to-pay levels which many jurisdictions use to determine decisions about the types of healthcare services and biomedical technologies they will invest in, and which are justified according to various repertoires of evaluation, such as the 'civic' and/or 'market' orders of worth (Boltanski & Thévenot 2006). Yet how these very same techniques were deployed within CANCERGEN signaled an altogether different, and more localized set of qualifications that had less to do with actual budgetary and accounting scenarios in healthcare delivery than with formalizing and streamlining the work of evaluating, valorizing, and commensurating across different choices at the stage of prioritization and design of clinical studies. As I describe in Chapter 5, this amounted to a process of *provisional economization* that produces a peculiar type of 'pre-public' economic number -- what I refer to as a form of *certified worth*, which corresponds to the coordinative dynamics of a circumscribed planning exercise rather than a broader coordination based on some legitimate public qualification.

The several features of CANCERGEN I have discussed throughout this section culminated in an understanding that both participatory and economizing activities within CANCERGEN demonstrated a certain pre-public quality, all the while possessing the possibility of being caught up in other future moments that might open them up to public legibility and the risks of contestation and critique which accompany it. There was often times also a reflexive move on the actors' behalf to stay below this public level and the qualifications the latter implies and so it is for this reason that I ultimately propose that in place of simpler distinctions between, for example, the public and the private, sociological analyses of coordinative action are well-suited by considering the wider *ecology of engagements and (e)valuations*, which I take up in several chapters including the conclusion to this thesis.

To conclude, then, the development of novel conceptual and theoretical notions which occur throughout this thesis are the result of a process of data gathering and analysis informed by the sociological approach known as grounded theory. They began not with my own interest in the given phenomena, but rather with a series of *sensitizing concepts* (Blumer 1954) which arose over the course of my research, including surveying the secondary literature as well as through my ethnographic

fieldwork methods of (participant-)observation and in-depth interviewing. Equipped with these concepts, I then traveled back to the literature in my own fields of sociology and STS as well as across the multiple moments within my fieldsite(s), to continue refining these notions. Returning to Becker's (1993) account of grounded theory one final time, he quips that '[i]ntuitions are great but they don't do much for us unless we follow them up with the detailed work that shows us what they really mean, what they can really account for' (35). Beginning with my own initial intuition about the difference between 'stakeholder engagement' versus processes of involving 'citizens' or the 'public,' I may have not have gotten very far if I had stopped at positing that these are simply different constituencies of actors who might be involved in medical and scientific decision making processes. However, probing deeper into these terms, and how they relate to particular modes of coordination and evaluation, I was able to say much more about CANCERGEN as well as about how it proves an instructive case for building out the analytic lexicons in social scientific analyses more broadly.

But before we can see the full utility of these notions, we first turn back to look at the emergence of comparative effectiveness research, its relationship to the wider ecology of practices in healthcare evaluative research, and how CANCERGEN came to stand in as a paradigmatic example of CER.

6. References

Abbott, A. (1988). *The System of Professions: An Essay on the Division of Expert Labor*. Chicago: University of Chicago Press.

Antal, A. B., Hutter, M., & Stark, D. (2015). *Moments of Valuation: Exploring Sites of Dissonance*. Oxford: Oxford University Press.

Asdal, K. (2018). 'Interested Methods' and 'Versions of Pragmatism.' *Science, Technology, & Human Values*, *43*(4), 748–755.

Baszanger, I., & Dodier, N. (2004). Ethnography: relating the part to the whole. In D. Silverman (Ed.), *Qualitative Research* (2nd ed., pp. 9–34). London: Sage.

Becker, H. S. (1993). How I learned what a crock was. *Journal of Contemporary Ethnography*, *22*(1), 28–35.

Blumer, H. (1954). What is Wrong with Social Theory? *American Sociological Review*, *19*(1), 3–10.

Boltanski, L., & Thévenot, L. (2006). *On Justification: Economies of Worth*. Princeton: Princeton University Press.

Cambrosio, A., Keating, P., Schlich, T., & Weisz, G. (2006). Regulatory objectivity and the generation and management of evidence in medicine. *Social Science & Medicine*, *63*(1), 189–199.

Carlson, J. J., Kim, D. D., Guzauskas, G. F., Bennette, C. S., Veenstra, D. L., Basu, A., … Ramsey, S. D. (2018). Integrating value of research into NCI Clinical Trials Cooperative Group research review and prioritization: A pilot study. *Cancer Medicine*, *7*(9), 4251–4260.

Carlson, J. J., Thariani, R., Roth, J., Gralow, J., Henry, N. L., Esmail, L., … Veenstra, D. L. (2013). Value-of-information analysis within a stakeholder-driven research prioritization process in a US setting: an application in cancer genomics. *Medical Decision Making*, *33*(4), 463–471.

Clarke, A. (2005). *Situational analysis: Grounded theory after the postmodern turn*. Thousand Oaks: Sage.

Deverka, P. A., Lavallee, D. C., Desai, P. J., Esmail, L. C., Ramsey, S. D., Veenstra, D. L., & Tunis, S. R. (2012). Stakeholder participation in comparative effectiveness research: defining a framework for effective engagement. *Journal of Comparative Effectiveness Research*, *1*(2), 181–194.

Deverka, P. A., Schully, S. D., Ishibe, N., Carlson, J. J., Freedman, A., Goddard, K. A., … Ramsey, S. D. (2012). Stakeholder assessment of the evidence for cancer genomic tests: insights from three case studies. *Genetics in Medicine*, *14*(7), 656–662.

Dussauge, I., Helgesson, C.-F., & Lee, F. (Eds.). (2015). *Value Practices in the Life Sciences and Medicine*. Oxford: Oxford University Press.

Eliasoph, N., Lo, J. Y.-C., & Glaser, V. (2018). Structured Ambiguity: How Meaning Emerges through the Faultlines of Institutional Logics. *Academy of Management Review*.

Fenstermacher, D. A., Wenham, R. M., Rollison, D. E., & Dalton, W. S. (2011). Implementing personalized medicine in a cancer center. *Cancer Journal*, *17*(6), 528–536.

Geiger, R. S., & Ribes, D. (2011). Trace Ethnography: Following Coordination through Documentary Practices. *HICSS 2011. Hawaii International Conference on System Sciences, IEEE*, 1–10.

Ginsburg, G. S., & Kuderer, N. M. (2012). Comparative Effectiveness Research, Genomics-Enabled Personalized Medicine, and Rapid Learning Health Care: A Common Bond. *Journal of Clinical Oncology*, *30*(34), 4233–4242.

Glaser, B. G., & Strauss, A. L. (1973). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.

Goffman, E. (1964). The Neglected Situation. *American Anthropologist*, *66*(6), 133–136. Retrieved from JSTOR.

Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is Originality in the Humanities and the Social Sciences? *American Sociological Review*, *69*(2), 190–212.

Hansen, M. P. (2016). Non-normative critique: Foucault and pragmatic sociology as tactical re-politicization. *European Journal of Social Theory*, *19*(1), 127–145.

Hine, C. (2007). Multi-sited ethnography as a middle range methodology for contemporary STS. *Science, Technology & Human Values*, *32*(6), 652–671.

Hoffman, A. S., Cambrosio, A., & Battista, R. (2016). Comparative Effectiveness Research in Health Technology Assessment. In A. Levy & B. Sobolev (Eds.), *Comparative Effectiveness Research in Health Services* (pp. 57–93).

Holden, M., Scerri, A., & Owens, C. (2013). More Publics, More Problems: The Productive Interface between the Pragmatic Sociology of Critique and Deweyan Pragmatism. *Contemporary Pragmatism; Leiden*, *10*(2), 1–24.

Kornberger, M., Justesen, L., Mouritsen, J., & Madsen, A. K. (Eds.). (2015). *Making Things Valuable*. Oxford: Oxford University Press.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford: Oxford University Press.

Latour, B., & Woolgar, S. (1986). *Laboratory Life: The Construction of Scientific Facts*. Princeton: Princeton University Press.

Lehoux, P., Daudelin, G., & Abelson, J. (2012). The unbearable lightness of citizens within public deliberation processes. *Social Science & Medicine*, *74*(12), 1843–1850.

Lezaun, J., Marres, N., & Tironi, M. (2017). Experiments in Participation. In U. Felt, R. Fouché, C. A. Miller, & L. Smith-Doerr (Eds.), *The Handbook of Science and Technology Studies* (pp. 195–220). Cambridge, MA: MIT Press.

Marcus, G. E. (1998). *Ethnography Through Thick and Thin*. Princeton: Princeton University Press.

Mol, A. (2002). *The body multiple: Ontology in medical practice*. Durham: Duke University Press.

Muniesa, F., & Callon, M. (2007). Economic experiments and the construction of markets. In F. M. & L. S. Donald MacKenzie (Ed.), *Do Economists Make Markets? On the Performativity of Economics* (pp. 163–189). Princeton: Princeton University Press.

Quéré, L. (1998). The still-neglected situation? *Reseaux. The French Journal of Communication*, *6*(2), 223–253.

Ribes, D. (2014). Ethnography of Scaling, or, How to a Fit a National Research Infrastructure in the Room. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 158–170.

Simonds, N. I., Khoury, M. J., Schully, S. D., Armstrong, K., Cohn, W. F., Fenstermacher, D. A., … Lyman, G. H. (2013). Comparative effectiveness research in cancer genomics and precision medicine: current landscape and future prospects. *Journal of the National Cancer Institute*, *105*(13), 929–936.

Star, S. L. (2010). Living grounded theory: Cognitive and emotional forms of pragmatism. In A. Bryant & K. Charmaz (Eds.), *The SAGE Handbook of Grounded Theory: Paperback Edition* (pp. 75–93). London: Sage.

Stark, D. (2009). *The sense of dissonance: Accounts of worth in economic life*. Princeton: Princeton University Press.

Stavo-Debauge, J. (2012, July 15). *La sociologie dite «pragmatique» et la philosophie pragmatiste, une rencontre tardive*. Presented at the Pourquoi le pragmatisme? L'intérêt du pragmatisme pour les sciences humaines et sociales., Villa Vigoni, Italy.

Thévenot, L. (2007). The plurality of cognitive formats and engagements moving between the familiar and the public. *European Journal of Social Theory*, *10*(3), 409–423.

Tsing, A. L. (2015). *The mushroom at the end of the world: on the possibility of life in capitalist ruins*. Princeton University Press.

Chapter 3

Comparative effectiveness research in health technology assessment

Abstract

Over the past several years, health technology assessment (HTA) and, more recently, comparative effectiveness research (CER) have become routinely deployed in various jurisdictions around the world. Despite some overlap in the ways that HTA and CER are used in making decisions about the use of health-care technologies, the relationship between these two sets of practices is still quite tenuous. There has been much debate about how these practices should be defined, what methodologies they should deploy in answering specific questions, and what standards should be used in adjudicating evidence about the utility of health-care interventions. In contrast to much of the health policy literature that attempts to prescribe what HTA and CER ought to be, the present chapter provides a historical–empirical approach to understanding the state of play of the HTA-CER nexus. In so doing, it explores issues presented by various research designs, including clinical trials, observational studies, systematic review, and meta-analysis, the advent of engagement practices, the emergent themes of patient-centeredness and personalization, and the problem of assessing the clinical and economic value of health technologies. After reviewing these issues, it moves onto examining one specific project in the United States, which serves to show how CER can be used in HTA, as well as how some of the more general problems discussed in the extant literature are dealt with in a more routine setting.

1. Introduction

The past several years have bore witness to a renewed sense of vigor for health technology assessment (HTA). In tandem with this resurgence has been the emergence of a new field of health care research called comparative effectiveness research (CER). This has been perhaps nowhere as noticeable as in the United States, where both terms were coined. These endeavors have now spread well beyond the American borders, both to the developed and the developing world, and have increased the scope and the intensity of discussions surrounding what HTA and CER are, what they ought to be doing, and how best to accomplish their respective goals. Unfortunately, as Luce and his colleagues from the

International Working Group for HTA Advancement (2010) recently pointed out, such discussions have resulted in a fair bit of confusion about how HTA and CER differ or relate to one another with regards to issues of regulation, decision-making, clinical practice, and the conduct of health care research more generally.

The purpose of this chapter is to move beyond programmatic approaches like those of Luce et al. by providing an empirical analysis of what CER and HTA *do*, as contrasted with what they *ought to be*. Such an empirical approach is more attuned to the situatedness of evaluative action; rather than focusing on *a priori* definitions of these phenomena, an empirical approach affords a more nuanced view that accounts for the still-developing models, methods, and applications that comprise the CER milieu. To foreground this argument, this chapter first examines the genesis of HTA and CER, which seeks to provide a better understanding of the distinct epistemological spaces that have been carved out in the literature by putting methodological and policy issues within HTA and CER in conversation. From there, the chapter moves onto an empirical analysis of a recent CER project in the U.S. in order to explore the issues, consequences, and stakes that are at play within this domain.

2. A brief history of HTA and CER

2.1 Health technology assessment

Both HTA and CER have firm roots in United States. For HTA, it was the Office of Technology Assessment (OTA) that first provided an infrastructure for conducting technology assessments, many of which at first focused on non-medical technologies -- for example, evaluations of mass transit, broadband communications, and automobile accidents (Banta and Jonsson 2009; Banta and Behney 2009). The establishment of such an office was a going concern as early as the late 1960s, and found a chief advocate in the Rep. Emilio Daddario -- then Chairman of the Science, Research and Development subcommittee of the House Committee on Science -- who sympathized with the calls of 'science advice' advocates in Washington for a strengthening of scientific expertise in government in order to better address the development and proliferation of new technologies at the level of national policy. Around this same time, a number of academic and popular publications began suggesting that the government 'was often failing to make informed choices about the use of science and technology,' culminating in the passing of the Technology Assessment Act of 1972 (Bimber 1996:27). The OTA was

thus established, with an early Chair of its Board stating: 'As an agency of, by, and for the legislative branch, the primary criterion of success for OTA is its ability to be of timely and useful service to the standing committees of the Congress' (quoted in Herdman and Jensen 1997:135).

Three years after its establishment and at the urging of key figures like Senator Edward Kennedy, the OTA initiated its health program whose mandate was to conduct what was at that time called 'medical technology assessment.' A first report, *Development of medical technology, opportunities for assessment*, was completed in 1976 and sought to relate the assessments of medical or health technologies to the broader field of technology assessment; a second report, *Assessing the efficacy and safety of medical technologies*, followed two years later and included sections on definitional and methodological issues, case studies exploring the safety and efficacy of medical technologies, an accounting of evaluative activities in the United States, as well as commentary on the availability of information on safety and efficacy. As two of the documents' authors have recently recalled, it was this latter report which

> might be the most important report done by OTA in terms [of] contributions to the development of HTA […] In many ways this report was ground-breaking. It pointed out the pervasive lack of accessible information on efficacy and safety, despite more-than-adequate methods of assessment. It also pointed to many problems that resulted from this lack, and the limited use of such information in clinical practice and policy making (Banta and Behney 2009: 29).

A third report, commenced that same year, was specifically attuned to the role of cost-effectiveness analyses of health technologies; a key feature of this document was its attention to the potential usefulness of cost-effectiveness studies in determining insurance coverage for health technologies. This was 'probably the first time this issue, which later became a key issue in the United States as well as other countries, was systematically raised and analyzed' (ibid.:31). Together, these three OTA reports laid much of the groundwork for the field of HTA and, in subsequent years, led to the establishment of formalized HTA bodies around the world: Sweden (1987), Canada (Quebec in 1988, and a later national program in 1990), the UK (1999), Germany (2004), and Denmark (2005), among many others. More recently, HTA agencies have also been formed in many developing countries and transitional economies, such as Brazil, Mexico, Malaysia, and China (Banta and Jonsson 2009).

In light of this proliferation, Lehoux (2005; 2006) points to two distinct phases in the evolution of HTA. Beginning in the mid-1980s, the scientific achievements of HTA came under global scrutiny,: a

wholesale push to standardize and refine those measures and methods used in HTA, such as the QALY (quality-adjusted life year), cost-effectiveness analyses, and the grading of evidence. It was also during this first phase that the International Society of Technology Assessment in Health Care (now HTAi) came into being, along with its official *International Journal of Technology Assessment in Health Care*, which 'yielded an enormous number of publications that have contributed to reinforcing, methodologically, the field's foundations' (2006:5). Despite the emphasis placed on methodological rigor during this first phase, however, there was in fact little in the way of 'theorizing HTA's goals and epistemological basis.' This issue became more evident in Phase II (ibid.:6) as defined by an increasing concern for how HTA products are actually taken up by their users (be they policymakers or otherwise), a concern that emerged as early as the mid-1990s (cf. Battista et al. 1994; 1999), and which now occupies center stage at annual HTA conferences around the world. The idea here is that different stakeholders -- such as regulators, insurers, or providers -- all expect different things out of HTAs, and so determining the ultimate impact of HTA reports requires understanding the initial intentions and intended audiences of a given assessment. Further, in many jurisdictions decision-making is dispersed and fragmented, so that targeting the message of an HTA is in itself no simple task. As Lehoux concludes: 'Phase II in HTA development consequently entails not only shaping an array of stakeholders' beliefs by providing them with scientific evidence about technology, but also understanding the regulatory mechanisms that may facilitate or impede the implementation of recommendations' (ibid.:7).

Despite the concerns that have surfaced in these two historical moments, the question remains: What *is* HTA? The International Network of Agencies for Health Technology Assessment (INAHTA), defines HTA as 'the systematic evaluation of properties, effects, and/or impacts of health care technology. It may address the direct, intended consequences of technologies as well as their indirect, unintended consequences. Its main purpose is to inform technology-related policymaking in health care [and] is conducted by interdisciplinary groups using explicit analytical frameworks drawing from a variety of methods' (Facey et al. 2006:27). Similarly, the European Network for Health Technology Assessment (EUnetHTA), defines HTA as 'a multidisciplinary process that summarises information about the medical, social, economic and ethical issues related to the use of a health technology in a systematic, transparent, unbiased, robust manner. Its aim is to inform the formulation of safe, effective, health policies that are patient focused and seek to achieve best value. Despite its policy goals, HTA

must always be firmly rooted in research and the scientific method' (EUnetHTA Website). Already, these two definitions list several key common features of HTA.

One feature is its stated purpose of informing the development of policy surrounding the use of health technologies. Goodman and Ahn (1999:98) have provided a rather inclusive list of policies which might be supported by HTA, including regulatory agencies deciding whether to permit commercial use of a technology, health care payers and providers deciding what technologies will be covered by benefit packages, clinicians and patients seeking to understand the proper use of technologies, hospital managers making decisions about which technologies to acquire, governments seeking to take up public health programs, technology developers exploring product development and marketing strategies, setting standards for the manufacture and use of technologies, and investors and companies considering industry transactions. Equally broad is the list of technologies that HTAs have sought to evaluate, which Lehoux (2006:46) has broken down into twelve categories; included here are screening tests, diagnostic tests and imaging devices, monitoring systems, implants, surgery and therapeutic devices, palliative technologies, drugs, health promotion technologies, occupational health technologies, technical aids, and information technologies. Some of these are 'hard' technologies -- for instance, drugs, diagnostic tests, surgery, and implants; conversely, health promotion, occupational health, and information technologies can at times be a bit less apparent to the naked eye, which of course does not mean their effects are felt any less.

Most definitions of HTA, including those put forth by INAHTA and EUnetHTA, include the aim of understanding the impacts of medical technology. INAHTA's definition is more general, highlighting the intended and unintended nature of such effects, while EUnetHTA's definition stresses particular types of effects: medical, social, economic, and ethical. There has been some level of debate within the HTA community about how such 'effects' are construed and how they are actually accounted for in the drafting of HTA products. A prime example of how this tension has manifested itself can be seen in an early example of HTA: OTA's first large single-technology assessment of the computed tomography (CT) scanner, which began in 1976 and was published two years later. As Banta and Perry (1997:432-3) state: 'Introduced to the market in 1972, this complex technology quickly became the prototype of the expensive tool in health care, the type of tool that was increasingly blamed for health care cost increases […] The CT scanner challenged the health care systems of many countries […] it was visible, exciting,

and expensive.' Elaborating further on the history of this assessment, Blume (2009:276) recalls his own correspondence with Banta about the drafting of this report: 'Banta and his staff appreciated that assessing the different consequences of health technologies required distinctive methodological approaches. Safety and efficacy, that could be established using epidemiological data and the results of controlled trials, would be the most straightforward. Moreover, these together with costs, were the aspects that principally concerned congress […] Reflecting Congressional concerns [the CT scanner report] focused on efficacy, safety, and financial costs alone. Efficacy, safety, and cost-effectiveness were the characteristics of medical technologies on which HTA gradually came to focus.'

Whence a third commonality between the two aforementioned definitions of HTA: both stress the use of specific methods, as well as a 'systematic' approach to analyzing health technologies. As can be seen in Blume's account, efficacy, safety, and cost-effectiveness are the cornerstone foci of HTA reports, each with its own definition and associated set of methods used to gauge its extent. Luce et al. (2010:261) posit that efficacy seeks to answer the question 'Can it work?' and explain that '[a] health care intervention is considered efficacious when there is evidence that the intervention provides the intended health benefit when administered to carefully selected patients according to prescribed criteria, often by experts in a research setting.' Safety can be understood as 'a judgment of the acceptability of the risks posed by the use of a technology' (OTA 1976). Meanwhile, [s]tudies of costs and related economic implications comprise a[nother] major group of methods used in HTA' (Goodman 2004:52). Despite these improvements in methodological rigor, the use of such information in the production of HTAs is not universally accepted, nor is it used in the same way across jurisdictions; this issue is addressed in greater detail in section 3.4 below.

This proliferation of economic analysis and methods is something of an irony in light of the fourth defining feature of HTA, namely its characterization as a multidisciplinary or interdisciplinary field/approach. Faulkner (2006) notes that the evolving nature of HTA during the 1990s, with its focus on effectiveness and cost-effectiveness of public health care, helped align the disciplines of clinical epidemiology (and its numerous subfields), public health, health economics, medical statistics, psychology, organizational analysis, general practitioners, medical and clinical science specialists, and even sociologists interested in medical and healthcare issues. He also remarks that the production of knowledge in HTA has been 'marked by uneasy and often unclearly-defined partnerships between these

disparate disciplines,' and which has at times signified a departure from many of the initial goals of HTA -- i.e., those that sought not only to evaluate efficacy and cost-effectiveness, but also the social and ethical elements of technologies. These latter concerns have largely been excluded from the historical development of the field, resulting in what Blume (2009) has called the 'narrowing down' of HTA.

One possible reason for this is the limited existence of methodologies that could systematically address ethical and social issues -- especially compared to the methods in clinical epidemiology and health economics deployed to generate data on efficacy and cost-effectiveness. Perhaps this was once true, although more recent work at the intersection of HTA and ELSI (ethical, legal, and social issues) indicates that there is in fact a multiplicity of ways of approaching these topics (cf. Bombard et al. 2011). A second explanation is the role of political pressures, which can be seen in the example of the influence the US Congress had on determining the scope of OTA's analyses. Other, more nuanced explanations have also been offered. Using the example of HTA's development in the UK, Faulkner (1997:201) suggests that 'it could be argued that if HTA practitioners are to focus on a scientific agenda built around generalizability, elimination of bias, and the representation of a form of aggregated public interest, this would preclude examination of substantive social and ethical issues.'

This point is reflected in Bombard et al.'s (2011) recent study on the integration of citizen involvement in HTA as a means of eliciting ethical and social values that ought to be reflected in the construction of technology assessments. Here, the authors report that the 14-person Citizens' Reference Panel on Health Technologies they recruited for their study ultimately settled on three core values -- 'universal access,' 'choice of options,' and 'quality care' -- and conclude that their data 'suggest[s] that decisions regarding the use and diffusion of health technologies should be guided by the principles of equitable access to technologies and the provision of choice to individuals, facilitated through trust-based relationships between patients and providers' (2011:6). When Panel members came to discuss how specific ethical and social values relate to discrete categories or types of technologies, they 'noted the need to identify relevant moral issues on a case-by-case basis,' yet the authors do not pursue this theme in any depth. While the study seeks to address those very social and ethical issues that have been excluded from most HTAs in the historical process of 'narrowing down,' in many ways the analysis indicates a general tendency for discussions about social and ethical values to default to issues of

economics vis-a-vis issues of access and equity, while more fundamental questions surrounding the unintended consequences of new technologies are left by the wayside.

Blume (2009) talks about a parallel 'narrowing down' within the field of bioethics during the same period in which HTA underwent such a transformation, where the issue of distributive justice and concern for the impact of biomedical advancements on community and individual welfare gave way to more simplistic notions of 'autonomy' and 'informed consent' (277). Yet for all the good will they portend, even these former preoccupations can still be said to have a certain bias towards what has been called the 'diffusion model' of technology (e.g. Latour 1987; Markussen and Olesen 2007), a phenomenon close in kind to that of technological determinism, which hones in on the assumed capabilities of technologies to reshape and reframe the 'material world of objects' without broader consideration of the dynamic interplay that exists between humans, technologies, practices, etc. Albeit an important criticism, further discussion of this issue is beyond the purview of the current analysis.

2.2 Comparative Effectiveness Research and Patient-Centered Outcomes Research

As with HTA, so too did the idea of CER originate in the United States; unlike HTA, however, the history of CER is a much more recent one. Why has the term experienced such a surge in use over the past several years -- with well under 50 publications per year using the combined term 'comparative effectiveness research' between 1960 and 2007, and which increased to almost 100 publications in 2008, nearly 250 in 2009, and upwards of 450 in 2010? While a more in-depth answer to this question is addressed elsewhere (e.g. Field 2016), the present chapter focuses briefly on two possible and mutually linked explanations. The first, covered in the context of the current historical review, is structural in nature and involves the drafting and passing of several pieces of legislation in the United States. The second, more pragmatic one touches on some of the key epistemological questions in contemporary biomedicine, and will form the basis of the following section in which the central themes within the emergent body of CER literature are reviewed.

With regard to the former, there has been a series of initiatives in the United States over the past decade that have focused on understanding the comparative effectiveness of medical interventions. The first of these was the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003, the largest overhaul of Medicare in its 38-year history, which primarily sought to encourage private insurance plans to participate in Medicare's coverage schemes and to form a new type of drug benefit

plan, although it also focused on restructuring outpatient drug coverage and provider payments (Wilensky 2004). The impetus for such a move was the growing recognition that, in spite of the increasing role of Medicare *viz.* the federal government in subsidizing prescription drug coverage for the nation's elderly and disabled as well as those with certain qualifying chronic illnesses, there was little in the way of a formal infrastructure for generating knowledge about the outcomes and effectiveness of these treatments (Smith 2007). Section 1013 of the MMA spoke most specifically to these concerns, and authorized the Agency for Healthcare Research and Quality (AHRQ) to produce systematic reviews, original research, new methods, and modes of communicating information about the 'outcomes, *comparative clinical effectiveness*, and appropriateness of health care items and services (including prescription drugs)' (MMA 2003, italics mine).

This was followed by the American Recovery and Reinvestment Act (ARRA, or the 'Stimulus Package'), which president Barack Obama signed into law on 17 February 2009. The ARRA came into being as a response to the U.S. financial crisis and injected about US$900 billion into the American economy through the financing of public works projects, ranging from local infrastructural initiatives like the building of roads to those aimed at much more fundamental levels of systemic change. Health care spending was a major focus of the Stimulus Package, with over US$150 billion apportioned for medical spending, something 'touted as a down payment on health care reform' in the U.S. (Manchikanti 2011:E253). Of this amount, $1.1 billion was set aside for 'comparative effectiveness research,' with the funds accordingly split between three major U.S. governmental agencies: $300 million to AHRQ; $400 million to the National Institutes of Health (NIH); and $400 to the Secretary of the Department of Health and Human Services (HHS), to be dispensed at her discretion. Aside from the financial windfall that the ARRA engendered for CER, however, the law also established the Federal Coordinating Council on Comparative Effectiveness Research (FCC-CER), whose mandate was 'to improve coordination of CER conducted and/or supported by the federal government and to help advise the Secretary [of HHS] on the allocation' of those moneys that HHS received (Birnbaum and Slutsky 2010:839).

Built into the ARRA legislation, however, was the demise of the very agency it created. This was done in order to transition from a body -- the FCC-CER -- whose members were employees of the federal government (half of whom were required to have clinical experience), to a new organization that, unlike both the FCC-CER and the OTA before it, would not be part of the government. The

FCC-CER ceased to exist on 23 March 2010, the very day that the Patient Protection and Affordable Care Act (PPACA) was enacted. In its stead, through the provisions of this overhaul to the American healthcare system, was the Patient Centered Outcomes Research Institute (PCORI) (PPACA, 2010). PCORI is led by a 21-member Board of Governors, including the Directors of AHRQ and the NIH, and an 18 member Methodology Committee consisting of experts in health care research methodologies from around the country. It is '[n]either an agency nor an establishment of the federal government,' but rather an independent, non-profit corporation that is arms-length from the government (Clancy and Collins 2010:1). The organization's mission is to 'assist patients, clinicians, purchasers, and policy-makers in making informed health decisions by advancing the quality and relevance of evidence' concerning ways in which a broad range of health conditions can be 'prevented, diagnosed, treated, monitored, and managed through research and evidence synthesis that *considers variations in patient subpopulations*' (ibid., italics mine). A striking feature of the transition from the FCC-CER to PCORI, aside from its relocation from within the federal government to a realm completely external to it, is also the terminology used to describe what each respective organization is doing. While the FCC-CER took as its focus 'comparative effectiveness research,' PCORI's mandate mentions very little about CER specifically, instead opting for the term 'patient-centered outcomes research' (PCOR). Such a move is not insignificant, as the wrangling over definitional issues has been a prominent part of CER/PCOR's recent history, perhaps part of the confusion Luce et al. (2010) have highlighted, and which is addressed further in the following section.

As briefly mentioned above, the idea of a 'comparative clinical effectiveness research' program in the United States was initiated in 2003 with the passing of the MMA, the first such legislative text to mandate CER. And yet the law itself says little in the way of what 'comparative clinical effectiveness research' actually *is* -- only insisting on its execution. In contrast, the legislative text of the ARRA in 2009 was much more directed, leading the FCC-CER to issue a definition of CER in its first report to the Congress:

> Comparative effectiveness research is the conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in "real world" settings. The purpose of this research is to improve health outcomes by developing and disseminating evidence-based information to patients, clinicians, and other decision-makers, responding to their expressed needs, about which interventions

> are most effective for which patients under specific circumstances' (FCC-CER 2009:18).

As stated by the authors of the FCC-CER report, this was the first time a standardized Federal definition of CER had been put forth. It was, however, neither the first nor the only definition to be published.

Chalkidou and Anderson (2009:a16), for example, in their review of international experiences with CER, highlight that as early as 2007 a number of both governmental and non-governmental bodies had sought to define CER as well; they highlight six such definitions from organizations as diverse as the American College of Physicians, the Institute of Medicine, the Medicare Payment Advisory Committee, the Congressional Budget Office, AHRQ, and the Pharmaceutical Research and Manufacturers of America. The authors conclude: 'Despite considerable overlap in the definitions, the table indicates significant differences, which were indicative of some of the debate over the role' of CER. While some of these debates are touched upon in the following section, the issue of 'value' (mentioned in only one of the six definitions), the role of primary versus secondary CER (i.e. the direct generation of evidence versus systematic reviews of existing bodies of evidence), and the option of focusing on the comparative effectiveness of health care interventions (i.e. screening, diagnostic, and therapeutic technologies) versus systematic issues (i.e. care delivery models) are mentioned here as being specific areas where such definitions tend to diverge from one another.

With the dissolution of the FCC-CER and the subsequent establishment of PCORI, similar definitional inconsistencies were encountered. Rather than simply having to pick up from where many of these previous definitions left off, PCORI's Board of Governors and Methodology Committee faced the challenge of defining something whose precise referent was entirely new. A 2011 document detailing the PCORI Methodology Committee's rationale for their definition of PCOR provides some sense of the shift from CER to PCOR. Bill 1796, which Senators Baucus and Conrad introduced to the U.S. Senate in October 2009, was the first bill suggesting the establishment of PCORI and the use of PCOR as an 'alternative term' for CER. The language of this bill was subsequently rolled into the Patient Protection and Affordable Care Act of 2010, which highlighted the need for a 'comparative clinical effectiveness research' enterprise and thereby established PCORI as a coordinating center for its conduct. And yet despite this mandate, the PCORI Methodology Committee claims that its members 'felt that from a patient's perspective, adopting a definition of PCOR that was synonymous with CER would not sufficiently describe the aspiration of this form of investigation,' arguing further: 'Not all

research that might be expected to help a patient make decisions or improve their experience in the healthcare system is explicitly comparative, and comparative evaluations do not necessarily incorporate the patient's voice, outcomes that matter to patients or comparisons that they value' (PCORI, 2011:2). Table 3.1 displays their definition of PCOR.

Table 3.1 PCORI Definition of Patient-Centered Outcomes Research

Patient-Centered Outcomes Research (PCOR) helps people make informed health care decisions and allows their voice to be heard in assessing the value of health care options. This research answers patient-focused questions:

1. "Given my personal characteristics, conditions and preferences, what should I expect will happen to me?"
2. "What are my options and what are the benefits and harms of those options?
3. "What can I do to improve the outcomes that are most important to me?"
4. "How can the health care system improve my chances of achieving the outcomes I prefer?"

To answer these questions, PCOR:

- Assesses the benefits and harms of preventive diagnostic, therapeutic, or health delivery system interventions to inform decision making, highlighting comparisons and outcomes that matter to people;
- Is inclusive of an individuals preferences, autonomy and needs, focusing on outcomes that people notice and care about such as survival, function, symptoms, and health-related quality of life;
- Incorporates a wide a variety of settings and diversity of participants to address individual differences and barriers to implementation and dissemination; and
- Investigates (or may investigate) optimizing outcomes while addressing burden to individuals, resources, and other stakeholder perspectives.

The extent to which the shift from CER to PCOR corresponds to more than a rhetorical or programmatic move, i.e. the extent to which it translates into distinctive, substantive and methodological developments and initiatives, and thus will lead to a radical redefinition of CER-related practices as contrasted to the mere broadening of some CER components, is still unclear at this stage.

3. State of Play: Exploring the HTA-CER Nexus

As discussed in the previous section, both HTA and CER each has its own unique history, and each has sought to account for the utility of health care interventions. Given that CER is a much newer

concept than HTA, the purpose of the current section is to outline the 'state of play' within CER vis-à-vis four central themes that have emerged out of the literature over the past several years and that characterize discussions surrounding the relationship between HTA and CER. Briefly, they are: issues and refinement of research methodologies; an emphasis on engagement practices; discussions about patient-centeredness and personalization; and a focus on value.

3.1 Issues and refinement of research methodologies in HTA and CER

In surveying the literature on methodological issues in HTA and CER, one quickly notices three main streams or approaches to evidence generation and synthesis: clinical trials, observational studies, and systematic reviews/meta-analyses. The evolution of CER in recent years has been marked by an increased focus on these three domains of research, and, in particular, by attempts to redesign clinical trial techniques to make them more adept at capturing the 'real world' significance and effectiveness of medical technologies.

3.1.1 Clinical trials

In a relatively early article on CER, Teutsch et al. (2005:128-9) remind us that while RCTs comparing active intervention with a placebo serve the needs of regulatory agencies in the United States 'and the desire to minimize the uncertainty surrounding efficacy assessments,' the past several years have seen increasing amounts of 'larger, longer-term RCTs using true health outcomes (such as mortality)' that have had a significant influence on clinical guidelines and thus clinical practice. Thus, in contradistinction to 'explanatory' placebo-controlled trials -- those that are designed to answer patho-biological questions under experimental (or ideal) condition' -- there is now a push to use more varied forms of the randomized trial with the goal of achieving a more 'cognitive approach to evidence-based policy making' that is believed to better answer a number of questions that are relevant to a broader variety of stakeholders (Tunis et al. 2010:1966).

Tunis and colleagues (2010) have recently outlined five different CER-relevant RCT methods, including pragmatic clinical trials, cluster RCTs, Bayesian/Adaptive trials, N-of-1 trials, and delayed-design or 'advance coverage' trials, each of which are briefly described here. Pragmatic clinical trials (PCTs) are significant for CER in that they specifically seek to understand the effectiveness -- rather than the efficacy -- of a treatment in routine clinical practice; their design 'reflects variations

between patients that occur in real clinical practice and aims to inform choices between treatments' (Roland and Torgerson 1998:285). Compared to placebo-controlled trials, which seek to eliminate bias outright through various features of their design and conduct, PCTs accept and capitalize upon these biases 'as part of physicians' and patients' responses to treatment and [are] included in the overall assessment' (ibid.). Luce et al. (2009) list some of the common elements of pragmatic trials, including 'clinically effective comparators, study patients with common comorbid conditions and diverse demographic characteristics, and providers from community settings.' (208). Cluster-randomized trials, like pragmatic trials, are prospective studies, but randomization is performed in 'clusters' rather than at the level of the individual patient; all patients in a given cluster are treated in the same way, and a cluster can involve patients being treated by individual physicians, by location, in group practices, by city/state, etc. The cluster-randomized study is then analyzed based on comparisons between the different clusters (Benson and Lyerly 2009).

Bayesian trials -- also known as adaptive trials -- are a third type of clinical trial. Their defining principle is that they estimate a priori distribution using prior information about a medical intervention. In contrast to the 'frequentist' school of statistics, which uses p-values to determine whether an intervention has a statistically significant impact at the trial's conclusion, the Bayesian approach to clinical trials uses

> formal, probabilistic statements of uncertainty based on the combination of all sources of information both from within and outside a study, [and] prespecifies how information from various sources will be combined and how the design will change while controlling the probability of false-positive and false-negative conclusions (Luce et al. 2009:207).

For CER, Bayesian methods are viewed as beneficial as they allow for competing interventions to be added or subtracted from the trial design while it is in progress such that 'the trial is comparing the alternatives most relevant to current practice,' thereby 'improv[ing] both the timelines and clinical relevance of trial results' (ibid.).

N-of-1 trials focus on a single patient rather than comparisons between groups of patients, and are often posited as one way around conducting expensive and lengthy clinical trials while still honing in on understanding precisely what treatments work in specific patients (i.e. Larson 2010). In an N-of-1 trial, the patient receives a 'series of pairs of treatment periods' during which active therapy is administered

for a period, followed by a second period where the patient is given either a placebo or an alternative treatment; random allocation dictates the order of treatments, and both patient and physician may be blinded as to which treatment is being administered at a given time in the sequence. As Guyatt and colleagues (1986) state, a key element of this type of trial is that

> a rapid improvement must occur when effective treatment is begun, and the improvement must regress quickly (but not permanently) when effective treatment is stopped. Selecting signs and symptoms that are particularly troubling or relevant to the individual patient provides one of the major advantages of the N-of-1 trial over conventional RCTs, in which tailoring of outcomes is generally sacrificed in favor of uniform end points that are applied to all study participants (890).

Finally, in delayed-design or 'advance coverage' trials all patients eventually receive the intervention under study, although some are administered the intervention from the study's outset, while for others the intervention is withheld at first for a designated period of time. Investigators are thus able to tell if patients benefit from shorter or longer durations of a given treatment, while at the same avoiding some of the ethical concerns that often arise during standard RCTs in which some patients are not offered an intervention that may potentially benefit them (Tunis et al. 2010:1969).

3.1.2 Observational studies

In the methodological literature, observational studies are often pitted against RCT data due to the issue of internal versus external validity. While many claim that the latter is the province of observational research while the former is most strongly allied with RCTs, this is not necessarily the case, as some view observational studies as complementary to RCTs. Black (1996), for instance, argues that experimentation may not always be necessary, appropriate, possible, or adequate. It is unnecessary, for example, when the magnitude of effect is so strong that the influence of confounding factors becomes virtually improbable. Experimentation is also often inappropriate for measuring infrequent adverse events, gauging the ability of an intervention to prevent rare events, and studying long-term outcomes, while at the same time random allocation may in fact reduce the effectiveness of the intervention under study. Furthermore, experimentation may be impossible when both physicians and patients resist enrollment in a clinical trial; there are also, in some instances, ethical objections, and political and legal obstacles to conducting RCTs (Black 1996:1216).

Black concludes that observational and experimental methods should thus be considered complementary: 'After all, experimental methods depend on observational ones to generate clinical uncertainty; generate hypotheses; identify the structures, processes, and outcomes that should be measured in a trial; and help to establish the appropriate sample size for a randomized trial' (ibid.:1218). Concato et al. (2010) put forward a similar argument, stating that 'critics of observational studies often "cherry pick" examples that support their a priori viewpoint,' whereas a more realistic conclusion is that 'the specific aspects of any particular study can be more important than the category of study design as randomized versus observational' (e18). With regard to CER specifically, the authors highlight that hierarchies of evidence that have been so dominant in the world of evidence-based medicine are being increasingly challenged due to the simplistic nature of their categorization, and so 'the right approach to a given CER study depends on the circumstances. At least one descriptive report suggests that both randomized trials and observational studies are being published as CER' (e21). Nonetheless, both bias -- i.e. systematic error -- and confounding -- i.e. mixing different effects together -- remain central to critiques of observational research (Dreyer et al. 2010:1819) and decreasing such systematic bias is 'perhaps the greatest challenge to using observational data sources for CER' (Tunis et al. 2010:1969).

There are four primary sources of observational data used in CER: administrative claims, electronic medical records, registries and other clinical cohorts, and case-control studies (Berger et al. 2009; Hlatky et al. 2012; Luce et al. 2009). Kim and Solomon (2011:1) state that administrative claims registries can be beneficial in the CER setting in that they provide detailed information on diagnosis, treatment, and disease activity, which in turn allows researchers to understand how different therapeutic interventions impact the regulation of a given condition in patients. Schneeweiss (2007) details additional benefits to using claims databases, especially in the context of post-marketing comparative effectiveness: the databases are relatively cheap to access; they represent 'real world' routine clinical practice; and are usually sufficiently large enough such that it shortens the amount of time to identify patients using a newly-marketed drug. Electronic medical records have similar advantages to administrative claims databases, especially in that they are relatively cheap to access and include information on diagnosis and treatment. However, EMRs have additional advantages such as providing clinical information that is timely and comprehensive, as well as often including physicians' notes and information on patient

symptoms and medical history that might capture certain nuances absent from administrative or insurance claims databases (Lau et al. 2011).

The data contained in clinical registries lay somewhere between data produced in RCTs and those included in electronic medical records and administrative or claims databases (Hlatky et al. 2012). Registries are similar to RCT data given the prospective and systematic nature of data collection used to capture patients' clinical, treatment, and outcomes-related information. The collection of these data is highly standardized compared to other sources of observational data, but registries also fundamentally differ from RCTs in that they rely neither on randomization to allocate interventions to different patient populations, nor do they employ the same restrictive inclusion and exclusion criteria in selecting research subjects. In this sense, registries are an important component of CER resulting from their provision of information that is much more reflective of 'real world' practices in heterogeneous populations. Finally, case-control studies identify individual patients who have experienced a given outcome, which are the 'cases', and others who have not experienced such an outcome, which are the 'controls.' In this type of study design, the cases and controls are compared in order to understand how being exposed to a specified intervention over a period of time results in either experiencing the outcome or not (Berger et al. 2009).

3.1.3 Systematic review and meta-analysis

Systematic review and meta-analysis is a third category of evidence used in CER. In contrast to RCTs and observational studies, which generate evidence, systematic reviews and meta-analyses focus on aggregating and synthesizing evidence that is already available. Systematic reviews are considered by many 'to be the best source of information for making clinical and health policy decisions. These research products rigorously summarize existing research studies so that health and health care decisions by practitioners, policymakers, and patients are more evidence based' (Whitlock et al. 2009). These summaries stipulate an a priori research question about a given clinical condition and proceed according to pre-selected methods to collect and analyze data from the clinical studies under consideration. A meta-analysis is more specific than a systematic review; while it also involves the pre-specification of questions and methods, it goes a step further by incorporating 'the statistical pooling of data across studies to generate a summary in the form of a pool of estimated effects' (Manchikanti et al. 2009, 930; O'Rourke 2007). Both meta-analyses and systematic reviews have both been key methods in health

technology assessment dating back to the mid-1980s, when a study of the use of streptokinase in treating acute myocardial infarction (AMI) indicated that the statistical analysis of a pooled group of smaller clinical studies rendered similar outcomes to those found in a single large study. The increasing popularity of these methods since that time can be explained by their ability to summarize a very large body of clinical information -- which is only growing -- and render different interventions comparable (Moreira 2007:182).

Specifically with regard to CER, such methods of evidence synthesis can be used for two primary purposes: to find gaps in evidence where further research can be conducted to understand the comparative effectiveness of interventions, as well as to use certain techniques to understand how different interventions compare in the absence of clinical studies that directly compare them (i.e. so-called 'head-to-head' trials) (Berlin and Soledad Cepeda 2012; Signorovitch et al. 2010). There has also recently emerged a new genre of systematic review, termed comparative effectiveness review, which has been defined as

> a unique type of systematic review which synthesizes the available scientific evidence on a specific topic. CERs expand the scope of a typical systematic review (which focuses on the effectiveness of a single intervention) by comparing the relative benefits and harms among a range of available treatments or interventions for a given condition (Manchikanti et al. 2009:931).

A significant methodological issue in this area of CER reviews, however, is how to compare interventions that have not been studied in a head-to-head trial -- something that is known as indirect comparison and/or mix treatment meta-analysis (also known as network meta-analysis). As with observational studies, here too there is much concern about introducing bias into CER systematic reviews, especially through comparing trials that have been conducted in different study populations. The Agency for Healthcare Research and Quality has published a methods guide for comparative effectiveness reviews, wherein it plainly states that, despite its recommendations:

> More studies are needed to determine when indirect comparisons are most likely to be valid. In the meantime, CER authors considering indirect analyses to assess harms should carefully consider whether assumptions underlying valid indirect comparisons are likely to be met, compare results of indirect comparisons with head-to-head data if available, and draw conclusions from indirect comparisons cautiously (Chou et al. 2012:124).

Moreover, in reference to the incorporation of observational data into CER reviews, the report also states that 'no grading system presently accounts for variations in potential risk of bias from different types of observational studies' but that reviewers should 'consider the question of value to the review with regard to each study design type' (Viswanathan et al. 2012:79).

More recently, some commentators have argued for the use of HTA reports and systematic reviews as a means of identifying evidentiary gaps and informing study designs in comparative effectiveness research. Tunis and Turkelson (2012) conducted a review of published literature relating to this issue and found a number of instances of this type of activity, including within AHRQ in the U.S. and NICE in the U.K. An interesting corollary to these exercises is that they occur in tandem with stakeholder engagement practices, wherein concerned parties -- including patients, clinicians, payers, hospital representatives, product manufacturers, and other stakeholders -- are brought in to give their expert perspectives as an adjunct to simply using evidence syntheses in identifying research gaps and prioritizing future research.

3.2 Engagement practices in HTA and CER

The increasing importance of engagement practices is in fact a second hallmark feature of CER. The notion of engagement may be viewed as a method in its own right, but where the previous section honed in on methods as they pertain to the generation and synthesis of clinical evidence, methods of engagement aim to move beyond the clinical realm and seek to account for the varying perspectives that exist in the multiplex world of health care decision making. There seems to be some consensus that including stakeholder input in CER can 'improve the relevance of research, increase its transparency, and accelerate its adoption into practice' (Concannon et al. 2012:985). These features of stakeholder engagement do not appear to be unique to CER, but are rather indicative of broader transformations in health care decision-making and policymaking more generally (i.e. Moreira 2012).

Engagement practices can be used to improve and legitimate policy directives, ensure transparency, increase public ownership of policy, assuage the so-called 'democratic deficit', and address the inherently complex nature of ethics that are intrinsic to priority-setting decisions (Bombard et al. 2011; Hodgetts et al. 2012). Yet identifying and defining what a 'stakeholder' is -- as well as what it means to 'engage' these actors, and to what ends -- is still very much an ongoing process in the CER milieu (i.e. Abelson et al. 2007). For example, in reviewing stakeholder representations in the field of genomics, Einsiedel (2009)

discusses the numerous domains in which the notion of 'stakeholder' has been deployed, including management, environment and resource management, policy, and international development, and offers two definitions of the term: Freeman's (1984) widely cited definition posits that a stakeholder is 'any group or individual who can affect or is affected by the achievement of the organization's objectives,' while the World Health Organization (2000) describes a stakeholder as 'any party to a transaction which has particular interests in its outcome' or who 'stands to win or lose by a line of policy.'

In a similar vein, Concannon et al. (2012:986) offer their own conceptualization of the term: 'An individual or group who is responsible for or affected by health- and healthcare-related decisions that can be informed by research evidence.' This latter definition is perhaps a more directed one in that it mentions health and healthcare as the specific domains within which stakeholders come together to deliberate, but it shares features of the two former definitions by highlighting the importance of understanding the organizational or policy impacts on specific constituencies, as well as the reciprocal impact that these constituencies can have in formulating specific organizational or policy objectives. Moreover, there is often some conceptual slippage between the use of the term 'stakeholder' and the use of the notion of 'the public' -- as in 'public engagement' or 'public participation' exercises. The present chapter contends that these two terms are not always mutually exclusive, but neither should they be used interchangeably; sometimes the 'public' can be a stakeholder group, while in other instances of stakeholder engagement a generalized 'public' is wholly absent from deliberative exercises. In reviewing the literature on stakeholder and public engagement, Deverka et al. (2012) maintain that the crucial difference is that 'stakeholders' have a distinct interest in a given health care issue and so their involvement in decision-making 'is seen as both rational and more likely to contribute to the quality and legitimacy or subsequent actions'; conversely, the terms 'public' or 'citizen' connote individuals who are unlikely to have a direct interest in a given issue (183).

3.2.1 Engagement in HTA

Notions of 'the public' as a stakeholder group are found in many instances of HTA, and are thus the focus of much of the recent literature on engagement practices in HTA. For example, Kreis and Schmidt (2012) review the public engagement processes used by HTA agencies in the three largest European economies -- France, Germany, and the United Kingdom -- all of which 'provide universal health care and face similar pressure to maximize efficiency and contain health care expenditures' (ibid.:91). Based

on their findings, they conclude that HTA organizations tend to involve the public in three main areas: public involvement in appraisal committees that seek to address a specific appraisal or coverage decision; a more indirect form of public involvement that invites the public to express views on more general principles of governance or decision making external to actual decision-making processes; as well as another indirect mode of involvement vis-à-vis lay involvement in executive bodies that influence certain substantive and operational decisions. The authors also highlight that, within these three areas of public involvement, there exist two primary modes of involvement: the first is when members of the public act as 'co-decision makers,' where they are afforded equal standing as experts and are often extended voting rights on a given decision; the second is when the public acts in a more 'advisory capacity', where public perspectives are brought to bear either on agenda-setting or else to provide personal experiences, views, and arguments that can influence certain decisions (ibid.:107).

In 2004, France established the National Authority for Health (HAS), which reports to the Department of Health and Parliament and is responsible for a variety of tasks in the health care sector. This includes the accreditation of health care organizations, certification of continuing professional development activities, developing guidelines for the treatment of chronic health conditions, and the provision of scientific advice to health authorities regarding assessments of drugs, medical devices, and diagnostic procedures covered under the national health insurance system. Final decisions regarding coverage of these technologies ultimately rests with the Department of Health, which is not mandated to follow HAS's advice. Assessments of clinical benefit within the HAS is performed by four committees. Two of these committees, the transparency committee and the committee for evaluation of medical devices and health technologies, do not involve any formalized processes for incorporating input from patients or from the public. The other two committees do formally involve the public; the committee for economic and public health evaluation has a 25 voting members of which two slots are reserved for patient and consumer representatives, while the committee for chronic diseases also reserves two of sixteen voting slots for consumers and patients (ibid.:101). Beyond these four committees, the HAS also incorporates public input into decisions beyond discrete clinical questions/conditions vis-à-vis focus groups and public consultations, where patient and user groups organize public stakeholder meetings that result in the drafting of a report summarizing these stakeholder views and that are made available to the committees that render an ultimate judgment on a

given topic. Unlike the examples of the UK/NICE and Germany/G-BA, France's HAS has no political mandate to involve the public in decision-making, although by doing so the intent is to ensure that 'patient needs are featured more on center stage in developing disease management guidance and accrediting hospitals and that patient information documents were tested for clarity and comprehensibility. Patient involvement on ethical matters also helped signal that HAS was pursuing a humane approach' (ibid.:104).

In Germany, the G-BA (Federal Joint Committee) was established in 2004, and although an independent entity, it works under the supervision of the Ministry of Health. Part of the G-BA's mandate is to make decisions about reimbursement policy for a variety of medical technologies, including drugs, diagnostics, therapeutic procedures, and medical devices and the committee issues legally binding directives to providers, sickness funds, and the insured population (ibid.:98). Up to five representatives from patient groups are permitted to participate in G-BA board meetings, along with five payer representatives and five representatives from the provider side. While the latter groups have voting rights, the patient representatives may only participate insofar as providing input into conversations, suggesting items they feel belong on the G-BA's agenda, as well as making recommendations for particular technology appraisals. To qualify as a patient representative, patients must be members of any number of groups that the German Ministry of Health recognizes as legitimate patient or consumer organizations, and it is up to those organizations to centrally coordinate the participation of patient representatives. Public engagement within the G-BA was provided for in the legislation that led to the Committee's creation, with the understanding that it was necessary for sickness fund users and their representatives to actively partake in decision-making activities as a 'necessary [part] in strengthen[ing] the role of personal responsibility' as well as encouraging 'more explicit and thorough justification of decisions. It also helped address some public perceptions according to which the G-BA was primarily a bureaucratic "black box" acting solely on the basis of statistics or numbers: the patient representative signaled that the people behind the numbers mattered' (ibid.:104).

Of the three jurisdictions discussed here, the United Kingdom has the most extensive history of public engagement in HTA, with the establishment in 1999 of the National Institute of Health and Clinical Excellence (NICE) that provides guidance to the British National Health Service on issues of clinical practice, health technologies, and public health. Public engagement in the drafting of technology

appraisals within the NICE framework involves the production of scoping documents that set out both the questions that an appraisal will seek to answer as well as identifying what organizations will be invited as official stakeholders (i.e. patient groups, manufacturers, and other relevant groups) to participate in the appraisal. Patient groups and care organizations that ultimately come to participate in an appraisal are invited to submit testimony on their experiences living with a given condition, as well as input about the technology or intervention that is under consideration, and they are also invited to provide a nomination of individuals who will attend appraisal committee meetings 'and contribute as expert witnesses, based on their own experiences' (ibid.:95).

The thirty-member committee that ultimately writes and reviews public input into a given appraisal also has three lay members that contribute equally to the proceedings, and this can include individual patients, carers, service users, or community members whom are invited to give a general patient perspective. The committee then offers up final draft guidance for a three-week period of public comment and, once closed, reviews any further commentary and ultimately submits final guidance to the NHS. These guidance documents are subsequently published in both long and short form. More recently, in 2002, NICE established its Citizens Council, which is intended to gauge the public's views and judgments on the clinical and economic elements of medical interventions in order to understand how citizens view what it means for an intervention to be good value, as well as the role of ethical and social norms in rendering these judgments. The perceived benefit of this Citizens Council is that it helps the governing board of NICE to 'get a better sense of the perspectives of ordinary people. Its reports directly informed the [Social Value Judgment documents] and generally required committees and the board to provide more explicit justification of decisions' (ibid.:104).

The importance of including ethical and social values in engagement practices within HTA -- seen most prominently in the UK, but also noticeable in France and more peripherally so in the German case -- is also visible in certain North American contexts. In Canada, the Ontario Health Technology Assessment Committee (OHTAC) along with the Medical Advisory Secretariat (MAS) strives to incorporate ethical and social values into its 'decision determinants' in drafting evidence-based recommendations. Between 2006 and 2010, several of the OHTAC committees took up a mandate to both incorporate these factors into its decision-making process as well as to engage with stakeholders about such issues. Interestingly, although the three previous European examples do indeed focus on

'ethical and social values', Kreis and Schmidt's (2012) review of European HTA agencies fails to make any conceptual distinction between 'ethical values' and 'social values.' Conversely, in the Ontario example, such a distinction has been made, with the former referring to 'questions related to the moral consequences of using the technology in preserving autonomy, integrity, dignity, etc.' and the latter indicating those 'questions that pertained to the distribution of resources, to commercial interests, to religious and cultural values, etc.' (Bombard et al. 2011:6).

Bombard and colleagues (2011) note that, despite this distinction being drawn in theory -- with 'ethical issues as principled, value-laden, normative assessments of HTA and the use of technologies' and social issues, which 'make the potential unintended consequences of their use explicit -- in practice such a distinction may not be tenable as they blend traditionally distinct disciplines of bioethics with social sciences and science technology studies' (Bombard et al. 2011:4). Notwithstanding this terminological issue, a public engagement committee within OHTAC-MAS ultimately ratified a plan for public consultation and identified three distinct stages where public input could be incorporated into the agencies' recommendations: early on, where the public could provide input on the questions a given HTA will seek to answer; at the stage of draft recommendation, where public input could help steer the recommendations that would result from the initial drafting of an HTA; as well as during the final stages of review, where the public would have a chance to comment on a draft HTA as it nears completion. Yet the methods for precisely how these ethical and social values should be incorporated into the HTA process are still very nascent, and the different categories of 'stakeholder' have not been formally outlined (ibid.).

The focus of Bombard and colleagues' own research is on integrating engagement mechanisms into Canadian HTA processes as a means of accounting for ethical and social values. Their work looks specifically at the role of the 'general public/citizens/taxpayer', which OHTAC's Public Engagement Subcommittee has operationalized as the 'friends or neighbors of patients, employers, members of local or cultural communities and citizens' (cited in Bombard et al. 2011:5). They also limit their analysis to investigating public input as it pertains to five technologies in the areas of diagnostic testing and population screening. In this limited context, they observe that across all five of the technologies around which citizens panels were invited to deliberate, the three key themes of 'universal access', 'choice' and 'quality care' were highlighted as 'core values that should be considered in the evaluation of health

technologies and ensuing recommendations' and that 'these core values suggest that decisions regarding the use and diffusion of health technologies should be guided by the principles of equitable access to technologies and the provision of choice to individuals, facilitated through trust-based relationships between patients and providers' (ibid.:8).

Although in many ways Bombard et al.'s conclusions fit with the themes of ethical and social judgments addressed within the aforementioned European HTA agencies, their research reveals that in the Canadian context -- at least as far as OHTAC-MAS is concerned -- there is still much uncertainty about precisely where public involvement is best incorporated in the lifecycle of an HTA document or recommendation, and at present there exists no formalized mechanisms for identifying different elements of 'the public' who will ultimately be invited to participate (Johnson et al. 2009). Such is not the case for other forms of 'stakeholder' engagement within the six-phase HTA cycle that OHTAC-MAS uses to draft HTA reports. For instance, the first phase of this process is called the 'vignette' stage, where the Medical Advisory Secretariat produces a description of the technology under consideration, and includes information on its expected magnitude of effect, pressures for technological diffusion, its potential to influence health care system efficiency and patient outcomes, as well as how it compares with other similar technologies (Levin et al. 2007:302). Relevant stakeholders who will be involved in later stages of the HTA are identified during this step, while public and stakeholder input is also sought in determining the outcomes that the HTA will assess to show the effectiveness of a given technology; this includes not only patients and 'the public,',but also partners in academia, expert physicians, and hospitals -- the latter of which are typically the purchasers of health technologies and thus those who will request that a specific HTA be carried out (ibid.).

A common feature shared amongst the four HTA organizations is that any involvement of stakeholders -- be it with patients, publics, or other actors in the health care system -- occurs as a post hoc exercise. That is, engagement is usually practiced at the point at which at least some clinical evidence has already been amassed regarding a certain technology, and so the input of actors within the health care field is geared towards understanding who will benefit from certain information, as in the case of stakeholder input into HTA prioritization, as well as ensuring that certain key decisions such as coverage of a technology under a national or provincial insurance scheme are transparent, accountable, and reflect the concerns of the many parties who are impacted by the existence and use of that technology.

3.2.2 Engagement in CER

While some of the same features found in HTA have carried over to the engagement practices used in the setting of CER and PCOR, there appears to be a broader trend within these areas to move engagement further 'upstream' (i.e. Greene, 2009). The discussion about engagement is a key area where one sees a substantial difference between the discourses of HTA and that of CER -- i.e. a practical rather than an *a priori* instantiation of difference between these two domains. This is perhaps best represented by the work of PCORI (see section two of this chapter for a history of this organization). In May 2012, PCORI's Board of Governors adopted its 'National Priorities for Research and Research Agenda,' which outlines five core areas of focus for the organization: assessment of prevention, diagnosis, and treatment options; improving healthcare systems; communication and dissemination research; addressing disparities; and accelerating patient-centered outcomes research and methodological research (PCORI 2011).

Of note here is that the first priority alludes to the generation of primary research evidence, and not simply syntheses of existing evidence, which are the primary loci for engagement practices in most HTA organizations. At the national level in the U.S., evidence syntheses have been typically carried out by the Agency for Healthcare Research and Quality, while primary evidence generation in the form of clinical trials and other forms of clinical research has generally rested with the National Institutes of Health. More recently, PCORI is coming to be viewed as a possible bridge between these two areas of research that have historically operated in separate silos (Clancy and Collins 2010). This upstream movement is something of a sea change, at least in the American health care research environment. Patient participation has surely played an important role in altering how medical research has been organized and carried out and how medical technologies have come to be regulated: one need only recall the AIDS epidemic and the activism it engendered during the 1980s (Epstein 1997) as a major example of this phenomenon, while cancer was itself a major rallying point for many other changes prior to this era (Carpenter 2010; Keating and Cambrosio 2011). And yet the establishment of PCORI through federal legislation signals a broad consensus that stakeholder and patient engagement have the potential to significantly enrich health care research in the United States.

There was in fact no specific legal mandate for the Institute to formally engage patients and other stakeholders, and yet PCORI's interest in stakeholder engagement was evident from its very inception.

This began with a series of public meetings of its governing board and methodology committee, which took place in various cities around the country over the course of several months, and where a separate 'Stakeholder Discussion Forum' was held in tandem with each meeting. The event announcement for the New York stakeholder forum contained the following passage:

> PCORI is committed to transparency and a rigorous stakeholder-driven process that emphasizes patient engagement. Patients will play a major role in PCORI's work by communicating what health care outcomes they value. PCORI will use a series of forums and formal public comment periods to increase awareness of its work and obtain public input and feedback prior to adoption of priorities, agendas, methodological standards, peer review process, or dissemination strategies (PCORI 2011).

During the discussion forum, members of the public -- including patients, physicians, and many other stakeholders -- were grouped at a number of tables along with one or two members of the PCORI Board of Governors and/or Methodology Committee, and were asked to discuss and report back the results pertaining to a number of issues: where one turns for reliable and trustworthy medical information, how one views the current state of medical information available, what questions one has about their own health or medical conditions, how one views the best way to engage patients and members of the public, etc. PCORI has also been active in recruiting stakeholders to participate as peer reviewers for PCORI funding applications, working in tandem with scientific reviewers to vet and prioritize submissions that best fit the Institute's mandate and stated interests.

This focus on engagement is further echoed in PCORI's 2011 Annual Report, which places 'the primacy of patient and stakeholder engagement' as the first of three areas of focus:

> This will remain our guiding principle […] We plan a wide array of multidirectional engagement initiatives involving patients, caregivers, clinicians, and other critical audiences […] We also plan to establish advisory groups of patients, caregivers and other stakeholders to help shape the direction of the organization's work in such areas as research networks, dissemination, clinical trials and observational studies, electronic health records, and building long-term PCOR. Driving home the value we place on engagement will be the inclusion of criteria requiring a robust, detailed engagement plan as part of all funding proposals (PCORI 2012:14).

A curious feature of PCORI's discourse around engagement practices is the conspicuous absence of a formalized rhetoric of ethical and social values, such as is seen in the aforementioned examples of HTA

organizations in the UK, Germany, France, and Canada. This may in part be reflective of PCORI's relatively young existence as an organization.. It may, however, also be the case that the roster of PCORI's 'stakeholders' -- while primarily focused on patients -- also includes a much broader range of actors within the health care system, resulting from a mandate that is much broader in scope than most other HTA organizations and is reflected in the 'upstream' movement addressed above.

3.3 Patient-Centeredness and Personalization in HTA and CER

The preceding discussion of engagement practices in HTA and CER is also closely tied to a third main theme that characterizes these two areas of research: a focus on patient-centeredness and personalization. Neumann (2012), for instance, states that patient-centeredness is closely linked to stakeholder engagement in that both '[acknowledge] the importance of giving a voice to affected parties and ensuring that clinical studies answer relevant questions' (586). Yet to talk about the notions of patient-centeredness and personalization first requires some level of conceptual clarity, especially since in many respects both terms address the issue of patient subpopulations.

There are in fact numerous ways to parse the idea of being attentive to 'patient subpopulations,' not to mention the various definitions of 'personalized medicine' that are deployed in the literature. One way is to consider 'holistic' approaches to medical care that include an accounting not just for patients' physiological characteristics, but also their personal, social, and emotional ones; this might include tailoring guidelines to account for the uniqueness of different patients' lifestyles, comorbidities, etc. (e.g. Braithwaite et al. 2007). A second way is to approach the issue in terms of the 'inclusion-and-difference paradigm,' which has consisted of a 'set of changes in research policies, ideologies, and practices' that 'reflects two substantive goals: the inclusion of members of various groups generally considered to have been underrepresented previously as subjects in clinical studies; and the measurement, within those studies, of differences across groups with regard to treatment effects, disease progression, or biological process' (Epstein 2007: 6). A third way of segmenting patient populations, and probably the most widely discussed in recent years, has come to be known as 'genomics and personalized medicine,' or GPM; this is a blanket term referring to the panoply of novel genomic tools that have proliferated through clinical practice over the past several years, and which use patients' genetic information -- rather than phenotype, as in the inclusion-and-difference paradigm -- to guide treatment and prevention strategies.

Interestingly, GPM seems to occupy a kind of third position that is concerned neither with a universalized notion of the patient, nor with patients as discrete individuals. Instead, it is more similar to what Epstein (2007) has called 'niche standardization,' which is 'a general way of transforming human populations into standardized objects available for scientific scrutiny, political administration, marketing, or other purposes that eschews both universalism and individual and instead standardizes at the level of the social group -- one standard for men, another for women; one standard for blacks, another for whites […] and so on' (135). However, GPM complicates Epstein's description because it focuses not on 'social groups' -- at least not initially -- but on '"stratified" medicine using genomics -- segmenting a patient population into subgroups based on hereditary risk of a disease occurrence, recurrence or likelihood of treatment response, or somatic changes in a tissue' (Conti et al. 2010:328). This is especially the case in the field of oncology, where the promises of 'personalized medicine' have in many ways been actualized and where a number of genomic tools have been introduced into clinical practice.

3.3.1 Challenges of personalized medicine in HTA

According to Becla and colleagues (2011), these kinds of personalized medicine technologies are thought to pose significant challenges to traditional HTA methods as they have existed until now. The authors draw comparisons between genomic tools and those difficulties encountered with the drafting of HTAs on orphan drugs for rare diseases. In the latter case, the dearth of large patient populations coupled with the limited nature of systems biology-based knowledge about human disease has precluded the conduct of large randomized controlled trials on these drugs; given the weight that RCT data have historically held in the drafting of evidence-based recommendations and coverage policy within HTAs, it has been difficult to make strong recommendations based on the available evidence. A similar argument is put forth by Merlin et al. (2012) with regard to 'codependent technologies' or companion diagnostics, where drugs are developed and marketed alongside tests that can determine the presence or absence of specific biomarkers in patients which help to predict whether patients will be more or less likely to respond to a given drug. This is especially true in the field of oncology, with codependent technologies now widely used in treating colorectal, lung, and breast cancers, and many others currently in the pipeline. Merlin and colleagues have evaluated the challenges that personalized medicine technologies have posed to HTA efforts in Australia given the limited evidence available upon which coverage and reimbursement decisions for codependent technologies can be made. The heightened focus on HTA for

evaluating genomic technologies, such as those currently used in the field of oncology, is not without reason. The present trend towards segmenting patient populations in health care research, and the 'market of limited application' that accompanies this shift, lead to an apparent paradox: at the same that time genomic technologies are challenging the very methodological basis of HTA, for many private payers HTA is coming to play a more important role in decision making regarding this specific group of technologies -- even more so, perhaps, than it does for other kinds of health care interventions (Trosman et al. 2011:22s).

One possible way of addressing the challenges of GPM technologies within HTA is to consider the broader translational pathway along which technologies travel, beginning with prototype discovery and design, through to preclinical development, clinical trials, and ultimately HTA and health services research. One of the primary limitations of current HTA practices, as noted by Becla et al. (2011), is the fact that HTA generally occurs at a relatively late stage in product development and deployment -- the so-called 'second gap in translation' -- where HTA is used as a tool for introducing interventions into clinical practice, but where the evidence required for regulatory approval may be insufficient for understanding 'real world' patient outcomes (2). Because of the complexity of new personalized medicine technologies, and the many challenges they pose for evidence development, the authors call for 'new models of HTA […] that can account for the specific type of evidence that is inherent to these novel therapies' (4). Here, the focus shifts from the second gap in translation to an earlier first gap, and the authors recommend Constructive Technology Assessment (CTA) as a possible means of evaluating new technologies. CTA begins prior to the technology being deployed in clinical practice and may in fact continue on to later stages of post-market economic analyses, which is at present a central feature of much HTA work internationally. According to Retel et al. (2012), 'CTA is based on the idea that during the course of technology development, choices are constantly being made about the form, function, and the use of that technology, and attempts to influence the development and diffusion of a new technology in a beneficial way' (442). Thus, by beginning to evaluate a technology's utility at an earlier stage, CTA is believed to produce knowledge that is both more timely and more relevant than 'traditional' later stage HTA activities; in this setting, patients can receive those interventions that are best suited to their particular disease or condition, while decision makers can have a wider range of cost-effectiveness data gleaned from different phases of a technology's deployment to refine further

implementation strategies going forward. Despite the increasing popularity of CTA methods, especially on the European continent, it is still quite far from the mainstream of health care research in North America.

3.3.2 Personalization and patient-centeredness in CER

Within the literature on CER, observers are nonetheless paying increasing attention to the need for moving evaluations 'upstream,' and for incorporating both patient preferences and individual biological characteristics into the CER paradigm. At the federal level, this is exemplified by the work of PCORI, as well as by the apportioning of funding for research focused on specific patient populations. PCORI's mandate explicitly specifies that the research it funds

> shall be designed […] to take into account the potential for differences in the effectiveness of health care treatments, services, and items as used with various subpopulations, such as racial and ethnic minorities, women, age, and groups of individuals with different co-morbidities, genetic and molecular sub-types, or quality of life preferences and include members of subpopulations as subjects in the research as feasible and appropriate' (PPACA 2010).

Interestingly, this covers each type of patient-centeredness and personalization that are discussed above, ranging from phenotypic to the genotypic characterizations of individuals.

On the ground, this has translated into more specific instances of patient-centeredness, such as the incorporation of patient reported outcomes (PROs) into CER. The US Food and Drug Administration has defined a PRO as 'any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else' (quoted in Basch et al. 2012:4250). According to Basch and colleagues, '[w]ithout including PROs, studies leave out essential information about the impact of interventions or health care processes on patients. The patient experience is at the center of most CER evaluations. Self-reports provide the most direct measure of the patient experience with disease and treatment' (ibid.:4251). In this sense, PROs are uniquely suited for capturing the patient's experience without the mediation of a clinician, as it is believed that clinicians often miss essential elements of patients' subjective experiences during the course of care. In an evidence guidance document devoted to incorporating PROs into CER in the area of adult oncology, Basch et al. (2012) used a multi-stakeholder approach to develop a set of fifteen recommendations broken into three main categories: selection of measures, implementation methods, and data analysis and

reporting, thus. elevating PROs to an essential position in CER and placing them on par with the essential clinical outcomes that are part of the standard reporting of clinical research results in the published literature. Especially significant here is that PROs seek to move beyond the accounting simply of survival measures, which have long been standard in clinical studies, towards a more holistic understanding of patients' experiences over the course of their disease and treatment trajectories.

3.4 A Renewed focus on value in HTA and CER

The themes of methodological expansion, public/patient/stakeholder engagement, and patient-centeredness and personalization are of obvious importance to the CER endeavor, as they have been in the refining of the HTA field before it. In many ways, the combination of these three phenomena signify an opening-up of research practices to incorporating phenomena that have often been left by the wayside in the orthodoxy of health care research over the past several decades. At the same time, there is a perhaps paradoxical closing-off to certain other features of health care research. An interesting element of the very legislation that institutionalized CER at the national level in the United States, and which subsequently led to the establishment of PCORI, is that it prohibits the development of certain societal average norms, such as measuring cost-per-quality adjusted life years (QALYs) in the evaluation of health care interventions. This is especially germane to the historical argument presented above in section two of this chapter and specifically to the notion of 'narrowing down' that according to Blume (2009) characterizes the history of HTA. That is, the field of HTA has undergone a process whereby many of its initial goals have given way to a wholesale focus on utilitarian strategies in priority setting, allocating health care resources, and maximizing the value of health care interventions (Brousselle and Lessard 2011:832).

3.4.1 Quantifying value in HTA

There are four main modes of economic evaluation that are deployed in HTA, each of which seeks to compare the costs and consequences of at least two health care interventions. Cost-benefit analysis (CBA) is concerned with issues of efficiency in allocating health care resources. In conducting a CBA, it is often the case that a dollar amount is applied to the number of life years an intervention is expected to gain for a patient, or other improvements in health and wellbeing it will engender; CBA does not only account for those benefits directly attributable to the intervention under consideration, but also

considers indirect costs and benefits that may derive from elsewhere in the health care system. Cost-effectiveness analysis (CEA), measures health effects in natural units -- such as years of life gained or average cases of disease that are avoided -- and compares the costs of achieving these objectives, placing a dollar amount on the cost per unit of effect; it is especially useful when different interventions result in similar effects. Cost-utility analysis (CUA) can be considered a special form of CEA that introduces a societal perspective on the health consequences of different interventions. The quality-adjusted life-year (QALY) is the most frequently deployed measure in the setting of a CUA, where a cost-per-QALY evaluation is used to understand the benefits of pursuing a given intervention or program over an alternative one. Of the four modes of economic analysis, CEA is thought to be most amenable to and practical for settings of decision-making because it uses a standard metric to compare interventions or other health care programs, allowing decision makers to visualize the different forms of worth that competing interventions or programs offer. Finally, cost-minimization analysis (CMA) is often the downstream result of either a CEA or CUA, analyzes whether alternative interventions or programs lead to sufficiently similar outcomes such that one strategy can be pursued essentially based on cost-savings alone (Goeree et al. 2011; Brousselle and Lessard 2011).

These four modes of economic evaluation are widely used by HTA agencies around the world, and their utility is often justified by the transparency that such modes bring to decision-making. In the United Kingdom, for example, the National Institute for Health and Clinical Excellence employs an explicit cost-per-QALY threshold of £20,000 to £30,000-per-QALY (approximately $30,000 to $45,000) in making coverage decisions (Le Pen 2009). At the same time, there has been much resistance to using strict economic thresholds in decision-making in other jurisdictions, an indication that economic evaluations are fraught with complexities and nuances that are not always easily addressed. Brousselle and Lessard (2011) provide a laundry list of such critiques, broken into four main areas. There are methodological issues, such as variation in utility scores used for calculating QALYs, the choice of respondents' influence on utility scores, the use of arbitrary discounting rates for costs and effects, and the imposition of arbitrary values to help standardize results which impose 'normative choice[s] about what should constitute good economic evaluation practice without resolving the methodological debates' (833). Second, there are contextual factors, such as the limited capability to generalize findings of economic analyses due to large variations in pricing and availability of health care resources between

jurisdictions, as well as the more general lack of influence that economic evaluations often have in health care decision-making due to decision-makers questioning if those data are relevant for their particular situation. Third, decision-maker factors exist, such as an inability for decision-makers to understand some of the more complex characteristics and terminology of economic evaluation, the time constraints placed upon decision-makers, and a disconnect between the outcomes of economic evaluations and what decision-makers actually consider when allocating resources. Finally, there are economic evaluation-related factors, such as concerns about the 'reliability, relevance, availability, and presentation' of studies (Brousselle and Lessard 2011:835), critiques of the QALY's failure to capture important health outcomes, conflict of interest issues in industry-sponsored economic evaluations, and the deployment of standardized assumptions that cannot account for the messy social realities (i.e. Harrington 2011).

In response to these criticisms and shortcomings, Brousselle and Lessard offer three possible alternative modes of economic evaluation. Cost-consequence analysis (CCA) is a non-aggregated presentation of costs and effects, which enables decision makers to view the predicted effects of an intervention, resource utilization, as well as humanistic outcomes. This method can also present the impact of an intervention across different populations (2011:836). Budget-impact analysis (BIA) measures the investment that the implementation of a health care intervention will require and estimates the affordability and necessary budgetary expenditures of the intervention in a real-world setting. Marginal analysis (MA) draws on the theme of stakeholder engagement and deploys a deliberative process for understanding the 'economic notions of margin (i.e. "the benefit gained or lost from adding or subtracting the next unit of resources for a given program") and opportunity cost' (ibid.). The authors, however, note that MA considers the effects of an intervention as secondary and often includes calculations based on the QALY, while also deriving comparisons between interventions from incremental cost-effectiveness ratios (ICERs). In this sense, marginal analysis is open to many of the same critiques that are waged against the four modes of economic analysis most frequently employed in the HTA milieu.

3.4.2 Quantifying value in CER

As with HTA, so too does the discourse surrounding CER involve discussions about health care costs, and here too there is much debate around and criticism of economic analyses. This is especially true in the United States, where there exists great skepticism about the role of cost-effectiveness analysis

and health care rationing. In the lead-up to the passing of the PPACA in 2010, these debates occupied much of the public discourse surrounding health care reform -- as is perhaps signified by the emergence of the term 'patient-centered outcomes research' as a more politically neutral alternative to 'comparative effectiveness research,' the latter of which many claimed would include cost-effectiveness analyses. Following from this, legislation stipulated that the permissible means of quantifying value would be primarily limited to clinical questions: 'The PCORI is specifically prohibited from adopting a QALY or similar threshold for establishing what types of care are cost effective, and Medicare coverage decisions cannot be made in a manner that "treats extending the life of an elderly, disabled, or terminally ill individual as of lower value" than in individual "who is younger, non-disabled, or not terminally ill"' (Harrington 2011:12). Interestingly, the Coverage Division within the Centers for Medicare and Medicaid Services, which is responsible for issuing national coverage decisions for Medicare, was prohibited from using cost or cost-effectiveness information to inform its judgments even before the 2010 health care reform legislation (Sullivan et al. 2009:s40). More generally, PCORI-funded research 'cannot be construed as mandates, guidelines, or recommendations for payment, coverage, or treatment. Coverage cannot be denied solely on the basis of CER. Evidence from CER can only be used as part of a larger process for making coverage decisions' (ibid.).

Neumann (2012) suggests that the rationale for approaching the issue of health care costs in the U.S. in such a manner is not without warrant, and reaches beyond the mere political pressure that elected officials face in ensuring Americans that their benefits will not be limited and their choices not curtailed: 'Research has revealed nonadherence to clinical guidelines, variation in practice patterns, preventable errors, and unnecessary hospitalizations. There is an overwhelming case for being smarter about how we finance and deliver care' (585). And yet, for all of the potential that a smarter, more patient-centered approach to health care research holds, Neumann also suggests that the focus on 'patient-centeredness' within the CER/PCOR research milieu is a somewhat paradoxical, if not problematic, situation: '[C]hanging the conversation to emphasize patients and stakeholders also has unhelpful consequences that few are willing to acknowledge. Focusing on patients' own preferences to the exclusion of considerations of societal resources will only compound our cost problems' (ibid.:586). Basu et al. (2011) offer a similarly skeptical position on CER, but move to empirically investigate the possible ways that

comparative effectiveness data might influence the pricing of medical interventions vis-à-vis producers' responses to new information on comparative effectiveness.

Counter to many of the popular beliefs about the promises of CER to control costs while improving health, Basu and colleagues argue that 'the impact of CER may vary with both the presence of insurance in markets and market power among producers or treatments,' further claiming that their 'analysis demonstrated instances in which CER may both increase spending *and* adversely impact overall health under plausible assumptions of how markets respond to the infusion of new quality information' (ibid.:10). The overall sentiment of their argument is that, in instances where cost-effectiveness data are precluded from aiding coverage determinations, the ability of CER to influence health care costs is variable and undetermined. They advocate for health economists to conduct continued 'CER of CER,' as they have done with their own quantitative framework, in order to help make more 'precise assessments of the value of public subsidies for technology assessments feasible and more common' (ibid.:11). This latter point is perhaps indicative of a broader trend in the CER milieu that appears to be gathering steam during the past several years, wherein economic evaluations take on something of a 'meta' quality and are being employed not only to evaluating the costs of health care interventions themselves, but also to understanding the value of a much broader range of phenomena, including the value of conducting research itself.

A significant example of this is the advent of value of information (VOI) (a.k.a. value of research) analysis. With roots in statistical decision theory dating back to the 1950s and 1960s, VOI was initially used in fields such as environmental risk analysis and engineering (Claxton and Sculpher, 2006; Wilson and Abrams, 2010). This set of approaches has gained traction in health care due to the fact that it 'values the additional information generated by further research in a way which is consistent with the objectives and the resource constraints of health care provision (the cost-effectiveness threshold)' (Claxton and Sculphur 2006:1056-7). In so doing, it allows observers to compare the costs of conducting further research on a given technology with the potential benefits of that research, while simultaneously accounting for the value of investing resources into the research study/studies under consideration. There are multiple levels at which value of information analyses can be performed in the setting of CER; given their complexity, this chapter offers only a high-level explanation of each level here. (However, for a more in depth discussion, see e.g. Claxton 1999; Hunink 2005; Meyers et al. 2011; Permonen et al.

2011). The first level is called the expected value of perfect information (EVPI), and produces a maximum value that a CER study is worth by calculating 'the probability that certain treatment choices based on current knowledge are suboptimal in terms of patient welfare and the potential welfare gain among patients that could be realized if these decisions can be remedied (with perfect information on comparative effectiveness)' (Helfand et al. 2011:191). The EVPI is calculated using an infinite sample size, and so at the second level -- the expected value of sample information (EVSI) -- an estimate is made about the value of a study's return given a specific sample size. (Wilson and Adams 2010:152). If the cost of doing the trial is less than the EVSI calculation, then it is generally regarded as research worthy of pursuing. At the third level, the expected value of partial perfect information (EVPPI) can be used to estimate the value of information gleaned from employing a certain set of parameters within a given study, e.g. mean survival for either a single intervention or for a group of interventions, such that knowledge about these parameters would lead to making the correct treatment decision; EVPPI provides an upper bound value for the returns that research on a given parameter set would result in.

Of these three levels, the EVSI is thought to be the most useful calculation for informing the prioritization of research studies 'because it allows comparisons against the cost of an actual trial […] If we are convinced that the costs of obtaining further information are equal to or lower than the expected VOI, it suggests that scarce resources should be allocated to collecting further evidence' (Husereau 2010:168). Such calculations are useful in real world instances of decision-making because of the comparisons that they facilitate, namely through the production of quantitative metrics of value that enables a side-by-side assessment of trials or interventions. This belief is echoed in a Brookings Institution report, published in 2009, which seeks to give a broad overview of the priorities, methods, and impact of CER. In the report, Garber and Meltzer (2009) discuss the potential role that value of information analysis can play in prioritizing CER: 'Prioritization can be viewed as an activity designed to maximize the value of a comparative effectiveness research effort […] The key issue in understanding how to maximize the value of specific CER efforts is to identify the mechanism by which such research will produce value […] [T]he chief purpose of CER is pragmatic: to produce information that changes clinical decisions for the better. This is a key element in the medical application of the principle of the "value of information"' (Garber and Meltzer 2009:18). Value of information and its attendant issues are examined in greater depth in the empirical example that follows. Suffice it to say that it is quite curious

VOI should gain so much traction in the U.S. at the very moment when economic analyses of health care interventions have become the target of much criticism and political wrangling. A possible explanation for this is that VOI as applied to prioritization provides an area in which U.S. health economists and those working in related disciplines are able to apply their methodological expertise while at the same time avoid charges of participating in 'rationing' of health care delivery by rationing research instead. The consequences of these activities remain to be seen, as VOI is still a very nascent approach, but it nonetheless appears to occupy an important place in the renewed focus on 'value' in the CER milieu.

4. An Empirical Approach to Understanding the HTA-CER Nexus

Of the US$1.1 billion in stimulus funding apportioned to build a CER enterprise, US$400 million was allocated to the National Institutes of Health, whose 'objective is to target funding to support scientific research opportunities that help support the goals of the Recovery Act. The projects support [the] Recovery Act by conducting CER that aims to enhance patient and clinician decision-making and to improve "real world" health outcomes for the Nation' (NIH 2010). A number of Grand Opportunities grants -- also called GO grants -- were developed out of these moneys, designed to help 'address large, specific biomedical and biobehavioral research endeavors that will benefit from significant 2-year funds […] The research supported by the "GO" grants program should have high short-term impact, and a high likelihood of enabling growth and investment in biomedical research and development, public health, and health care delivery' (NCI 2009). These GO grants were awarded for research in a variety of health care settings, and focusing on many different disease categories. Reflecting the 2009 Institute of Medicine report that prominently featured issues in oncology in its top 100 priority areas for CER in the U.S., several projects received funding from GO grants to study issues in the personalization, prevention, and care of cancer patients (IOM 2009). Seven of these GO-grantee institutions were selected to form a consortium focused on conducting CER specifically on genomics and personalized medicine technologies in oncology. This final section draws from a single one of the aforementioned GO consortium projects as an empirical example for understanding how CER can be used in HTA: the Center for Comparative Effectiveness Research in Cancer Genomics, or CANCERGEN.

4.1 CANCERGEN: CER meets the Cooperative Clinical Trials Program

Based in Seattle, Washington, CANCERGEN is in fact a large research center consisting of several different organizations: the Fred Hutchinson Cancer Research Center in Seattle; the University of Washington, also in Seattle; the Center for Medical Technology Policy (CMTP), in Baltimore; and SWOG (formerly the Southwest Oncology Group), a cooperative oncology group whose headquarters are presently located at the University of Michigan. The Center's overarching mission is to facilitate the movement of promising cancer genomics technologies through the innovation pipeline in such a way that the technologies can be used to improve the effectiveness as well as the cost-effectiveness of clinical cancer care. The deliverables of this initiative incorporate the varying interests of its collaborating institutions: to develop a comprehensive process to prioritize emerging cancer genomics technologies that can be evaluated through the SWOG clinical trials network; to integrate CER within RCTs to facilitate rapid design and implementation of CER trials in priority areas; to design a proof-of-principle comparative effectiveness study alongside a SWOG-run clinical trial; and to develop analytic policy models to support clinical trial design and early-stage technology assessment (Ramsey et al. 2009).

To deliver on these four aims, each of the institutions participating in CANCERGEN plays a unique role that speaks to the consortium's multifaceted approach to HTA and CER. In general, all of the consortium's activities are coordinated through the Fred Hutchinson Cancer Research Center, where the staff is responsible for data linkage and management, as well as statistical analyses of trial data. SWOG is the hub for clinical trial design and trial data management, statistical analysis of trial results, and patient reported outcomes. The University of Washington contingent of CANCERGEN is based in the school's Pharmaceutical Outcomes Research and Policy Program, and its work in the consortium involves simulation modeling, cost effectiveness analysis, ethical evaluation, as well as study designs for Coverage with Evidence Development (CED) trials. Finally, the Center for Medical Technology Policy (CMTP) is responsible for recruiting the CANCERGEN External Stakeholder Advisory Group, conducting horizon scanning and priority setting exercises, designing pragmatic clinical trials, as well as providing guidance on CED trials. Based on this very brief description, one can easily recognize several issues at the HTA-CER nexus that have been covered in the previous 'state of play' section. This speaks to the way that institutions are building up innovative infrastructures that are co-evolving with the

broader landscape of outcomes research in the U.S, combining many different elements into new constellations that are, in turn, changing the very nature of this domain.

Since it was initially funded in 2009, the work of CANCERGEN has led to numerous publications in the clinical and policy literatures. A primary focus of these articles has been on the processes -- especially in the area of stakeholder engagement -- that the consortium has developed to help achieve its goals of being 'sustainable, collaborative, [and] multidisciplinary.' One such publications argues that CER should be understood as involving both evidence synthesis and evidence generation practices in comparing different approaches to diagnosing, treating, and preventing disease, as well as comparing alternative systems of health care delivery: 'While a broad range of study designs such as systematic reviews, observational studies and randomized controlled trials are included in CER, what distinguishes this field of research is its particular purpose' (Deverka et al. 2012:181). The authors then highlight the 'critical distinction' between the Institute of Medicine's definition of CER -- to help 'consumers, clinicians, purchasers, and policy-makers to make informed decisions that will improve healthcare at both the individual and population levels' -- and that of the 'traditional construct of health research' that they claim 'often places stakeholders as passive audiences for research results, rather than directly informing priority areas and study design' (ibid.). They seek to draw attention to the fact that scientists and researchers generally design clinical studies without necessarily understanding whether the results of those studies will be applicable and relevant to decision-makers in the health care system, and who must make tough choices about how to treat patients, what expenses to cover under a given benefit plan, how to best regulate the use of medical technologies, etc.

Using this as their point of departure, the authors move on to argue that 'stakeholder engagement is a fundamental, and perhaps defining, aspect of CER' (ibid.). This claim is striking in that it subtly suggests CER be understood as something that is much more of a 'work-in-progress' rather than a stabilized phenomenon. That is, stakeholder engagement is *a* fundamental aspect -- and not necessarily the only one -- that characterizes CER; moreover, it is also a *perhaps defining* aspect of CER, i.e. not necessarily so. By defining CER in these terms, the CANCERGEN consortium nonetheless makes certain kinds of actions both possible and justifiable (Berger and Luckmann 1966; Boltanski and Thevenot 2006) by enacting CER as something that is inextricably linked with the input of stakeholders. Stakeholder engagement is instrumental in bridging the gap between research and decision-making vis-à-vis the

active involvement of various stakeholders across the many different phases of designing and conducting research on medical technologies. As such, CANCERGEN is an instructive case study for understanding one particular instantiation of how CER is used in HTA, affording a privileged view of the development of specific networks that exemplify the multiplex nature of the CER-HTA nexus.

Beginning with SWOG's role is useful, as it immediately gets to some of the methodological issues that are at stake in CER. SWOG is comprised of over 4000 researchers at 500 different institutions, including 22 NCI-designated cancer centers as well as institutions in several other countries, and it operates approximately 100 clinical trials with around 30,000 subjects at any given time, with an average of 5000 research subjects enrolled in its studies each year. In recent years, SWOG has begun to reevaluate the way it functions, initiating a quality assessment and improvement initiative in 2007 to better understand its processes and systems, what can be improved, and how to ensure the continued generation of quality data from the clinical trials that SWOG runs (JOP 2008). As such, it is in fact one of the largest and most prolific of the National Cancer Institute Cooperative Groups currently in existence in the United States. The NCI Cooperative Group Program was initiated in 1976, and established four central objectives, which continue to guide the Program today: 1) to make state-of-the-art cancer management available to cancer patients in the community; 2) to involve a wider segment of the community in clinical research than is possible through the existing cooperative group programs; 3) to enhance recruitment of patients from community hospitals into appropriate protocols; and 4) to evaluate the transfer of new patient care technology to the community (SWOG 2012).

Despite the important role that cooperative groups have played in conducting landmark clinical research studies in oncology, the past several years have seen a change in tone regarding how the NCI's Cooperative Group Program should operate. Notably, in 2010, the U.S. Institute of Medicine convened a consensus conference and subsequently published a landmark report detailing the promises and pitfalls of the Program. The report, 'A National Cancer Clinical Trials System for the 21st Century: Reinvigorating the NCI Cooperative Group Program,' celebrates the Cooperative Group Program's work of producing knowledge on cancer care and prevention, which has led to decreased incidences of cancer and improved survival and quality of life for cancer patients (Nass et al. 2010). And yet even in its preface, the report's timbre begins to shift:

> Publicly sponsored trials fill an important information void by conducting head-to-head comparisons of different treatment regimens, combining

> treatments, and investigating whether drugs approved for the treatment of one type of cancer can be used to effectively treat other types of cancer, all of which are far less likely to be pursued by pharmaceutical companies. However, the NCI Cooperative Group Program is falling short of its full potential to improve the quality of care that cancer patients receive. An accumulation of problems is hampering progress, just at a time when new knowledge about the genetic and molecular underpinnings of cancer has created opportunities for designing trials with new, targeted anticancer agents. Increasingly, biomarkers (predictors of a response to a particular therapeutic intervention) can be used to select which treatment strategy is most likely to benefit individual patients. (Nass et al. 2010:xi).

The report goes on to highlight two central flaws that have dramatically impacted the Program's work: the first is the 'lengthy and redundant' set of processes required to initiate clinical trials in the Cooperative Group framework, which 'results in frustration and a perception that stakeholders are working at cross-purposes' (ibid.). The report also highlights the 'terrible waste of human and financial resources' resulting from the inefficient processes for prioritizing clinical trials and for choosing certain trials that are likely to be successful, claiming that only a bit more than half of all NCI-sponsored trials actually reach completion and have their results published (ibid.). The second pitfall is a 20% reduction in funding for NCI-sponsored trials that has occurred since 2002, while at the same time 'new knowledge of the molecular changes underpinning cancer and the use of predictive biomarkers in cancer therapy not only increase the potential impact of trials but also add to their complexity and cost' (Nass et al. 2010:2). This has led observers including Scott Ramsey, the Principal Investigator of CANCERGEN, to characterize the contemporary American clinical trials enterprise as being 'inefficient and cumbersome' (Scoggins and Ramsey 2010:1371).

By approaching CANCERGEN in the broader context of the IOM report, it becomes evident that there are many convergences between the clinical trials enterprise and the current discourses of HTA and CER. It is no accident that SWOG is the 'laboratory' of choice for testing the CANCERGEN 'experiment,' and CANCERGEN's very first deliverable was to collaborate on a major clinical trial run by SWOG. This trial, called the RxPONDER study (short for Treatment for Positive Node, Endocrine Responsive Breast Cancer Study, SWOG protocol number S1007), is a phase III, prospective, randomized controlled study testing the clinical utility of Genomic Health's 21-gene assay Onco*type* DX test in 4,000 women with newly-diagnosed hormone receptor-positive, Her2-negative breast cancer with

1-3 positive lymph nodes. Onco*type* DX uses a proprietary 'recurrence score' that stratifies breast cancer patients into one of three groups -- either low risk, intermediate risk, or high risk -- for developing recurrent disease following surgical excision of the breast tumor. One of the novel features of the Onco*type* DX test is the inclusion of this intermediate risk group. Up until the test was made commercially available, recurrence was generally only understood dichotomously, i.e. a patient either had a low risk or a high risk of developing recurrent disease. The inclusion of this new intermediate risk group has led to a significant level of uncertainty surrounding how patients who fall into that category should be treated, and so the RxPONDER trial is especially interested in answering clinical questions about the intermediate risk patient population.

Previous studies of the Onco*type* DX test have shown the test's ability to provide accurate prognostic information in patients with *lymph node-negative* breast cancer treated with tamoxifen (Ray 2009). The central clinical problem the RxPONDER study addresses is the impact of chemotherapy on patients with *node-positive* breast cancer who have either a 'low' or 'moderate' Onco*type* DX recurrence score, as well as to understand the optimal recurrence score cutpoint for prescribing chemotherapy. The guiding hypothesis of the trial is that the recurrence score will predict the benefit of chemotherapy for patients who fall under the aforementioned diagnostic criteria, and that the benefit of chemotherapy will increase as recurrence scores increase. If the hypothesis is correct, this could potentially impact the treatment strategies for a large subset of patients, as approximately one third of women present with node-positive malignancy at the time of initial diagnosis.

4.2 The External Stakeholder Advisory Group

The work of CANCERGEN and its emphasis on stakeholder engagement is in many ways set up to address the critiques of randomized controlled trials, which have been discussed previously. What matters to clinicians and clinical researchers is not always the same as what matters to patients, and the questions that are often asked in a typical RCT may not produce answers that are relevant to stakeholders elsewhere in the healthcare system -- be it patients, regulators, insurers, or consumer advocates. Furthermore, in addition to the ongoing theme of stakeholder engagement, CER is in many ways a departure from the orthodoxy of the clinical trials enterprise due to the increasing efforts aimed at expanding, refining, and experimenting with novel approaches to clinical and outcomes research. What, then, is to be made of the fact that CANCERGEN is explicitly bound up with the U.S. clinical

trials cooperative group program which has until this point in history been almost exclusively concerned with conducting randomized controlled trials? A recently published article from CANCERGEN claims that the Cooperative Group Program is in fact an ideal platform for conducting CER trials in oncology, and specifically in personalized medicine:

> Prospective randomized, controlled trials create the most compelling evidence to change clinical practice, but they are artificial constructs using highly selected patient populations. To address this issue, clinical trials groups could implement naturalistic or pragmatic trials […] The Clinical Trials Cooperative Group Program supported by the National Cancer Institute offers and infrastructure for high quality prospective comparative effectiveness research on genomic tests for cancer. The program's mission is to promote and support clinical trials of new cancer treatments and diagnostics, explore methods of cancer prevention and early detection, and study quality-of-life issues and rehabilitation during and after treatment (Ramsey et al. 2011:2265).

The 'proof-of-principle' RxPONDER study is the first major move towards incorporating CER into the Clinical Trials Cooperative Group Program, with Group Chair, Dr. Lawrence Baker, stating: 'We're proud to be the first among the ten NCI cooperative groups to embrace -- and to be funded for -- comparative effectiveness research as part of our mission' (University of Michigan 2010). The recent attempts to incorporate CER objectives into SWOG clinical trials have, however, been somewhat controversial, with CANCERGEN participants feeling resistance from certain groups within the cancer clinical trials community, so that much needs to be done in order to get 'buy in' from many trialists (i.e. Hoffman et al. 2010).

When CANCERGEN was initially funded by the Grand Opportunities grant in 2009, it swiftly began interfacing with SWOG and its RxPONDER by way of convening an External Stakeholder Advisory Group (ESAG), a 13-member panel consisting of representatives from both public and private insurers, pharmaceutical and device industries, clinicians, and patient/consumer advocates. There was great incentive for CANCERGEN to begin working towards achieving its objectives as quickly as possible. In convening the ESAG, there was an immediate push to evaluate what was, at that time, the ongoing design of the RxPONDER study. The primary outcomes of interest in the study have already been mentioned: to examine the impact of chemotherapy in patients with node-positive breast cancer with intermediate or low Onco*type* DX recurrence score, as well as to determine the optimal recurrence score

cutpoint for recommending chemotherapy. The role of the ESAG, then, was to help decide secondary endpoints of interest.

One of the most significant changes that was made to this study as a result of the ESAG input involves an element of the study's inclusion criteria, which initially stipulated that in order to enroll in the trial, women would have to undergo complete axillary nodal dissection. It had been thought that, by removing axillary nodes following a positive sentinel lymph node biopsy, women would show either increased overall survival rates or else a reduction in axillary recurrences in breast cancer following lumpectomy and radiation to the breast. Axillary dissection is often accompanied by significant side effects, including 'pain, restricted range of motion, discomfort, and lymphedema in the affected arm' (ASCO 2011). The ESAG members were solicited for their input in the RxPONDER design and expressed reservations about this stipulation at the time. Interestingly, approximately one month following the ESAG's decision to alter the inclusion criteria, an abstract presented at the 2010 annual meeting of the American Society of Clinical Oncology reported data from a large randomized controlled trial showing that after six years of follow up, there were in fact no differences in treatment outcomes between a group of women that received only a sentinel lymph node biopsy and those that received SLNB plus follow-up axillary node dissection; the findings were subsequently published in the Journal of the American Medical Association (Giuliano et al. 2011). The revised eligibility requirements stipulate that axillary node evaluation may be performed per the standard of care at the institution at which the patient is undergoing evaluation, but that it is not required for study entry.

4.3 Patient-centeredness and economic evaluation in CANCERGEN

An additional secondary objective included in the RxPONDER study, and largely influenced by ESAG input, is the use of patient reported outcomes (PRO) instruments to study patient attitudes. Members of the ESAG were especially interested in understanding pre-testing patient preferences as well as background information on their knowledge of the disease and general education levels that could possibly impact one's decision to enroll in the study and consent to being randomized (Ramsey et al. 2013:5). The study plans to recruit 1000 patients in the pre-randomization phase, both prior to being tested with Onco*type* DX as well as following the receipt of recurrence score results, and uses five different survey instruments to study issues of patient anxiety, concern about cancer recurrence, patient health state preferences, concerns prior to testing, as well as a questionnaire about patients' choice for

adjuvant treatment to better understand patient treatment choice. An additional 500 patients will be administered survey instruments following treatment randomization; here, the same five tools will be used to study post-randomization issues including anxiety, patient health state preferences, fatigue and cognitive impairments, survivor concerns, and other issues they might have following randomization in the trial.

As discussed in the state of play section above, this gives researchers a picture of patients' experiences of using the Onco*type* DX test without the mediation of a clinician and provides valuable data capturing patient preferences at various points throughout the trial, beginning with pre-testing and continuing until after they receive treatment. The incorporation of PROs also fills a significant gap in the literature, where at present there exist 'no studies gathering information round the time of the decision of whether or not to take the test, and following the results, the factors that most influence patient decisions about treatment' (ibid.:4). By generating more and better data on patient preferences in the setting of Onco*type* DX, it is believed that clinicians will be better able to navigate potential obstacles to using the test, to accepting its results, and to using those results to inform treatment decisions.

This use of PROs and the generation of information on patient preferences is tied to a third element of CER that has been integrated into the RxPONDER study: the use of trial data to perform economic evaluations of Onco*type* DX. This falls in line with certain understandings of CER that claim economic analyses as a central feature of its conduct. This position is supported by several CANCERGEN-affiliated researchers: 'Cost-effectiveness analysis is one […] technique falling under the umbrella of comparative effectiveness research in which both effectiveness and costs are considered: which intervention works better, by how much, and how much more (or less) does it cost?' (Thariani et al. 2012:4). The question of costs is becoming increasingly important in the area of genomics and personalized medicine; the Onco*type* DX test has a default list price of US$4,175.00 (elsewhere listed at approximately $3,500.00) and is widely used, with the already expensive price tag further compounded by uncertainty surrounding how much more useful this test is compared to current standards of practice for rendering diagnostic and prognostic information.

As with the inclusion of PROs, cost-effectiveness analysis was also not a feature of the initial trial design and only became a going concern after the consolidation of CANCERGEN and the convening of the ESAG. The trial is being used as a platform to understand the cost effectiveness of using the

Onco*type* DX risk score in node-positive patients versus usual care. The trial, however, was not designed to directly compare management of breast cancer with the Onco*type* DX test versus management without the test, which significantly complicates many of the standard ways of gauging the economic impact of medical technologies, such as those covered in section 3.4.1 above. Given these limitations, the ESAG suggested that certain data elements be clarified that could facilitate cost-effectiveness estimations of the 21-gene assay. This resulted in the addition of modeling techniques as well as the collection of disease-free survival (DFS) data in addition to other types of survival data that were initially included as primary endpoints for the RxPONDER study. Other sources of data were also subsequently included following ESAG deliberations, including health state utilities from a survey instrument utilized as part of the PRO collection process and collection of health insurance records to further aid in estimations of cost-effectiveness (Ramsey et al. 2011:6).

Interestingly, despite the efforts to perform cost-effectiveness analyses within the context of CANCERGEN, and even in light of the standard role that CEA plays in assessing the value of medical technologies -- as is witnessed throughout the history of HTA -- Goddard and colleagues (2012) claim that genomics can in fact be a challenging area for the application of these economic methods. The authors highlight that there is, in the first place, an overall lack of comparative effectiveness data on genomic technologies, which contravenes most efforts at gauging the comparative value of these tools and so analysts must be cautious in their assessments of uncertainty within this area. Moreover, there is also difficulty in measuring the value that doctors and patients place on knowing genetic information, thereby complicating the process of incorporating issues surrounding genomics into policy decisions (Goddard et al. 2012:638). In many ways, the very fact that the aforementioned modifications were made to the RxPONDER trial following ESAG deliberations attests to the difficulty of performing economic analyses at the CER-genomics nexus and shows some of the work-arounds that are necessary in order to obtain certain types of information.

4.4 Value of information in CANCERGEN

Aside from looking at the value of medical technologies themselves, the advent of CER has also been accompanied by an increased interest in using the aforementioned value of information (VOI) or value of research (VOR) approaches. Recall that the central purpose of this stream of research is to help make decisions about what technologies are most amenable to CER evaluation and how to best design CER

trials such 'that additional research reduces our uncertainty about which intervention to use in clinical practice' (Goddard et al. 2012:639). Accordingly, an important part of the CANCERGEN project has thus been to develop and improve VOI/VOR methods for use at the CER-genomics nexus. The use of VOI/VOR in CER studies of genomics may be especially useful given the rapid pace with which advancements in genomics are taking place and the consequential need for methods to help prioritize comparative studies that are often very expensive to run. Within CANCERGEN, the value of research studies applied to the RxPONDER trial sought to understand the benefits and harms of Onco*type* DX compared to the best available current therapeutic options for treating women with node-positive breast cancer; it also attempts to comprehend the societal value that this research study offers.

The primary question the VOI model asked was whether using Onco*type* DX in node-positive patients to select out this specific population who would benefit from adjuvant chemotherapy was better than the current standard of practice which suggests that all patients in this group receive adjuvant chemotherapy. The model was primarily concerned with the expected value of sample information (EVSI), which as stated earlier, provides a monetary return value for a given trial based on a specific sample size and overall trial design. As VOI is a modeling technique that can incorporate many different types of data, including both clinical and economic information, the model for the RxPONDER trial drew on previous clinical trial results as inputs for the analysis of patient outcomes; this included using data from an older trial of node-positive breast cancer patients to estimate overall 10-year survival rates in light of the paucity of such data specific to the population under study in the RxPONDER trial, as well as drawing on data from a previous Onco*type* DX study to estimate disease-free survival rates. The VOI model also looked to the extant literature for input on quality of life issues, such as cost and utility of the genomic test; the model derived quality-adjusted life-year calculations from life expectancy data that had been modeled elsewhere, while the costs associated with treatment and disease recurrences were based on insurance claims data. Patient preferences were also considered, but very little data is available regarding how patients with node-positive breast cancer make decisions about whether or not to receive adjuvant chemotherapy, and so the RxPONDER model incorporated a fairly wide range as to whether would choose to receive or forgo chemotherapy (Wong et al. 2012:1120).

The expected value of sample information for the RxPONDER study was ultimately calculated at three different levels of willingness-to-pay, which is a standard metric that accounts for how much a

given society is willing to spend on an additional year of life. The model's output indicated that the study would be worth anywhere from $450 million for the lowest threshold up to $1.05 billion for the uppermost WTP level; this accounts for the 20,600 cases of node-positive, HR-positive, Her2-negative breast cancer diagnosed each year and multiplied over a 10 year span. An interesting element of the CANCERGEN/RxPONDER VOR model was that it not only evaluated the EVSI of the final trial design, but also attempted to understand the study's value both prior to involving stakeholders in its design as well as after the aforementioned additional data elements were incorporated. Here, the original protocol design was evaluated as being worth between $400 and $960 million at the same three willingness-to-pay thresholds, and the additional parameters that resulted from the ESAG deliberations added between $50 and $100 million to the trial's overall value (Wong et al. 2012:1122).

In any of these calculations, when compared with the projected NIH trial budget of $27 million that it would cost to run the study, it becomes evident that there is 'a projected return on the investment of 17 to 39 times the NIH trial cost, suggesting the study is a good investment in research resources. These findings were driven by 1) the high level of uncertainty in outcomes based on current evidence, 2) the high incidence of breast cancer, and 3) the severity of clinical and economic outcomes in node-positive disease' (ibid.). However, the External Stakeholder Advisory Group's impact on the value of the study is not as straightforward as it might first appear. As the authors point out, the ESAG input that led to the inclusion of endpoints relating to cost, utilities, and patient preferences in fact added very little to the value of the study when considered individually; it is only when the three elements were considered together, and combined with survival outcomes, that the increased value became visible (ibid.). Moreover, the fact that survival measures -- rather than issues of cost -- have a greater influence on patient outcomes, and are also the major driver of value within the RxPONDER study, suggests that CER studies on health care interventions that are of greater concern to stakeholders may be of greater value. This is because there is likely a greater willingness to pay for positive outcomes from a health care intervention that a diverse group of stakeholders considers to be important.

While CANCERGEN's VOI analysis of the RxPONDER study is informative with regard to gauging the value of this individual study, it falls somewhat short of many descriptions and applications of VOI/VOR, especially the use of these techniques in designing and prioritizing potential trials of health care interventions. This is at least in part due to the fact that the RxPONDER study design had already

been finalized and had moved on to begin recruiting patients, and so the VOI analysis was a post-hoc exercise aimed more at understanding the possible utility of using such approaches in the setting of multiple-stakeholder engagement, which is a central feature of the CANCERGEN project. Further work has been initiated by CANCERGEN to expand the use of VOI as part of a broader strategy for prioritizing a second clinical trial to be carried out by SWOG. In this second deployment of these techniques, calculations actually demonstrate to stakeholders the comparative worth not only of different study designs, but also differences in value between competing disease areas and medical interventions; this includes evaluations of EGFR and ERCC1 mutation testing in the setting of non-small cell lung cancer, as well as the use of certain tumor markers for detecting recurrence of breast cancer following primary therapy (Thariani et al. 2012). In this way, VOI offers a formalized and transparent methodology that can help justify why certain CER studies are pursued and why expenditures might be allotted to a given disease area and/or medical intervention over its competing alternatives. Moreover, it exemplifies a major point of convergence between HTA and CER, which are both moving towards adopting VOI for health care priority setting in numerous domains and jurisdictions (i.e. Claxton and Sculpher 2006).

5. Conclusion

The purpose of this chapter has been to explore the relationship between the set of practices called HTA and those that have more recently been termed CER. It first began by exploring the unique histories of HTA and CER, each of which is rooted primarily in the U.S. but which have since proliferated to much of the developed and even developing worlds. From there, it moved on to detail the current 'state of play' at the HTA-CER nexus, which consists of four central features: a refinement of research methods, the deployment of engagement practices, a focus on patient-centeredness and personalization, as well as a renewed focus on value. Finally, in the last section, the chapter moved to show how each of these four themes has been incorporated and expressed in a major CER study in the United States. In reviewing the progression of the fields of HTA and CER, it becomes evident that despite the many attempts to differentiate these two fields from one another, the more important question is in fact not what they are -- and how they differ -- but rather what they do, how they function, and what the consequences of these strategies might be. As can be seen from the empirical

example of the CANCERGEN project, there are many novel elements that the advent of CER has introduced into the health care research milieu in the United States.

These shifts hold great promise for improving further upon many of the issues that have at times been recalcitrant to moving research practices and patient care forward. The willingness to go beyond the randomized controlled trial, to focus on population subsets whether based on phenotype or genotype, to engage stakeholders in the prioritization and design of CER studies, and to reconsider what it means for trials and technologies to be 'valuable', are all perceived to be necessary elements of a new paradigm of health care research. And yet CER is still a very young field and so what observers presently perceive as a major shift may in fact be only the beginning of a much longer process of significant change within biomedical research and health care evaluation. Consistent attention to how these practices are changing and in which contexts is essential to identify the pressures spurring these changes , and to obtain a better picture of the twin fields of HTA and CER, a broader understanding of how these two sets of practices function in contemporary society, and an indication of how they might continue to change, thereby affecting the activities not only of doctors and patients, but also of a variety of other stakeholders that are constitutive of today's highly differentiated health care system.

6. References

Abelson, J., Giacomini, M., Lehoux, P., & Gauvin, F.-P. (2007). Bringing 'the public' into health technology assessment and coverage policy decisions: from principles to practice. *Health Policy*, *82*(1), 37–50.

Banta, D., & Behney, C. J. (2009). Office of Technology Assessment health program. *International Journal of Technology Assessment in Health Care*, *25*(S1), 28–32.

Banta, D., & Jonsson, E. (2009). History of HTA: introduction. *International Journal of Technology Assessment in Health Care*, *25*(S1), 1–6.

Banta, H. D., & Perry, S. (1997). A history of ISTAHC: A personal perspective on its first 10 years. *International Journal of Technology Assessment in Health Care*, *13*(03), 430–453.

Basch, E., Abernethy, A. P., Mullins, C. D., Reeve, B. B., Smith, M. L., Coons, S. J., … Eppard, W. (2012). Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. *Journal of Clinical Oncology*, *30*(34), 4249–4255.

Basu, A., Jena, A. B., & Philipson, T. J. (2011). The impact of comparative effectiveness research on health and health care spending. *Journal of Health Economics*, *30*(4), 695–706.

Becla, L., Lunshof, J. E., Gurwitz, D., Westerhoff, H. V., Lange, B. M., & Brand, A. (2011). Health technology assessment in the era of personalized health care. *International Journal of Technology Assessment in Health Care*, *27*(2), 118–126.

Benson III, A., & Lyerly, K. (2009). Improving medical decisions through comparative effectiveness research: cancer as a case study. Friends of Cancer Research.

Berger, M. L., Mamdani, M., Atkins, D., & Johnson, M. L. (2009). Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part I. *Value in Health*, *12*(8), 1044–1052.

Berger, P., & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. Garden City: Anchor Books.

Berlin, J. A., & Cepeda, M. S. (2012). Some methodological points to consider when performing systematic reviews in comparative effectiveness research. *Clinical Trials*, *9*(1), 27–34.

Bimber, B. A. (1996). *The politics of expertise in Congress: The rise and fall of the Office of Technology Assessment*. Albany: SUNY Press.

Birnbaum, H., & Slutsky, J. R. (2010). Guiding Comparative Effectiveness Research -- A US Perspective. *Pharmacoeconomics*, *28*(10), 839–842.

Black, N. (1996). Why we need observational studies to evaluate the effectiveness of health care. *BMJ: British Medical Journal*, *312*(7040), 1215-8.

Blume, S. S. (2009). Assessing health technologies in a changing world. *International Journal of Technology Assessment in Health Care*, *25*(S1), 276–280.

Boltanski, L., & Thévenot, L. (2006). *On Justification: Economies of Worth*. Princeton: Princeton University Press.

Bombard, Y., Abelson, J., Simeonov, D., & Gauvin, F.-P. (2011). Eliciting ethical and social values in health technology assessment: A participatory approach. *Social Science & Medicine*, *73*(1), 135–144.

Braithwaite, R. S., Concato, J., Chang, C. C., Roberts, M. S., & Justice, A. C. (2007). A framework for tailoring clinical guidelines to comorbidity at the point of care. *Archives of Internal Medicine*, *167*(21), 2361–2365.

Brousselle, A., & Lessard, C. (2011). Economic evaluation to inform health care decision-making: promise, pitfalls and a proposal for an alternative path. *Social Science & Medicine*, *72*(6), 832–839.

Carpenter, D. (2010). *Reputation and power: organizational image and pharmaceutical regulation at the FDA*. Princeton: Princeton University Press.

Chalkidou, K., & Anderson, G. (2009). *Comparative effectiveness research: international experiences and implications for the United States*. Washi AcademyHealth.

Chou, R., Aronson, N., Atkins, D., Ismaila, A. S., Santaguida, P., Smith, D. H., … Moher, D. (2008). Assessing harms when comparing medical interventions. In *Methods guide for effectiveness and comparative effectiveness reviews. AHRQ Publication No. 10 (12)-EHC063-EF*. Rockville: Agency for Healthcare Research and Quality, 112–129.

Clancy, C., & Collins, F. S. (2010). Patient-Centered Outcomes Research Institute: the intersection of science and health care. *Science Translational Medicine*, *2*(37), 37cm18..

Claxton, K. (1999). Bayesian approaches to the value of information: implications for the regulation of new pharmaceuticals. *Health Economics*, *8*(3), 269–274.

Claxton, K. P., & Sculpher, M. J. (2006). Using value of information analysis to prioritise health research: some lessons from recent UK experience. *PharmacoEconomics*, *24*(11), 1055–1068.

Concannon, T. W., Meissner, P., Grunbaum, J. A., McElwee, N., Guise, J.-M., Santa, J., … Leslie, L. K. (2012). A new taxonomy for stakeholder engagement in patient-centered outcomes research. *Journal of General Internal Medicine*, *27*(8), 985–991.

Concato, J., Lawler, E. V., Lew, R. A., Gaziano, J. M., Aslan, M., & Huang, G. D. (2010). Observational methods in comparative effectiveness research. *The American Journal of Medicine*, *123*(12), e16–e23.

Conti, R., Veenstra, D. L., Armstrong, K., Lesko, L. J., & Grosse, S. D. (2010). Personalized medicine and genomics: challenges and opportunities in assessing effectiveness, cost-effectiveness, and future research priorities. *Medical Decision Making*, *30*(3), 328–340.

Deverka, P. A., Lavallee, D. C., Desai, P. J., Esmail, L. C., Ramsey, S. D., Veenstra, D. L., & Tunis, S. R. (2012). Stakeholder participation in comparative effectiveness research: defining a framework for effective engagement. *Journal of Comparative Effectiveness Research*, *1*(2), 181–194.

Dreyer, N. A., Tunis, S. R., Berger, M., Ollendorf, D., Mattox, P., & Gliklich, R. (2010). Why observational studies should be among the tools used in comparative effectiveness research. *Health Affairs*, *29*(10), 1818–1825.

Einsiedel, E. (2009). Stakeholder representation in genomics. In P. A. Atkinson, P. E. Glasner, & M. Lock (Eds.), *Handbook of genetics and society: Mapping the new genomic era* (pp. 187-202). Abingdon: Routledge.

Epstein, S. (1996). *Impure science: AIDS, activism, and the politics of knowledge*. Berkeley: University of California Press.

------. (2007). *Inclusion: The politics of difference in medical research*. Chicago: The University of Chicago Press.

EUnetHTA. (2013). Common questions – health technology assessment. What is health technology assessment? http://www.eunethta.eu/faq/Category%201-0#t287n73. Accessed 15 Jan 2013.

Facey, K., Topfer, L.-A., & Chan, L. (2006). INAHTA Health Technology Assessment (HTA) glossary. Stockholm: International Network of Agencies for Health Technology Assessment. Available at: http://inahta.episerverhotell.net/upload/HTA_resources/Edu_INAHTA_glossary_July_2006_final.pdf

Faulkner, A. (1997). "Strange bedfellows" in the laboratory of the NHS? An analysis of the new science of health technology assessment in the United Kingdom. *Sociology of Health & Illness*, 19(19b), 183–207.

------. (2006). In the sociomedical laboratory of citizen health: exploring science, technology, governance and engagement in prostate cancer detection in the UK. Cardiff School of Social Sciences Working Paper Series, No. 74. Cardiff University.

FCC-CER [Federal Coordinating Council for Comparative Effectiveness Research]. (2009). *Report to the President and the Congress, June 30, 2009*. Washington, DC: The Council.

Field, R. (2016). Comparative Effectiveness Research and Health Reform in the USA. In A. Levy & B. Sobolev (Eds.), *Comparative Effectiveness Research in Health Services* (pp. 41–56). Boston: Springer US.

Freeman, R. E. (1984). *Strategic Management: A Stakeholder Approach*. Boston: Pitman.

Garber, A. M., & Meltzer, D. O. (2009). Setting priorities for comparative effectiveness research. In *Implementing Comparative Effectiveness Research: Priorities, Methods and Impact* (pp. 15–33). Washington, DC: Brookings Institution.

Giuliano, A. E., Hunt, K. K., Ballman, K. V., Beitsch, P. D., Whitworth, P. W., Blumencranz, P. W., … Morrow, M. (2011). Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: a randomized clinical trial. *JAMA*, *305*(6), 569–575.

Goddard, K. A., Knaus, W. A., Whitlock, E., Lyman, G. H., Feigelson, H. S., Schully, S. D., … Khoury, M. J. (2012). Building the evidence base for decision making in cancer genomic medicine using comparative effectiveness research. *Genetics in Medicine*, *14*(7), 633–642.

Goeree, R., He, J., O'Reilly, D., Tarride, J.-E., Xie, F., Lim, M., & Burke, N. (2011). Transferability of health technology assessments and economic evaluations: a systematic review of approaches for assessment and application. *ClinicoEconomics and Outcomes Research*, *3*, 89–104.

Goodman, C. S. (2004). Introduction to health technology assessment. *The Lewin Group. Virginia, USA*.

Goodman, C. S., & Ahn, R. (1999). Methodological approaches of health technology assessment. *International Journal of Medical Informatics*, *56*(1), 97–105.

Greene, J. A. (2009). Swimming upstream: comparative effectiveness research in the US. *Pharmacoeconomics*, *27*(12), 979–982.

Guyatt, G., Sackett, D., Taylor, D. W., Ghong, J., Roberts, R., & Pugsley, S. (1986). Determining optimal therapy—randomized trials in individual patients. *New England Journal of Medicine*, *314*(14), 889–892.

Harrington, S. E. (2011, January 1). Incentivizing comparative effectiveness research. Ewing Marion Kauffman Foundation Research Paper. Available at: http://www.kauffman.org/uploadedFiles/Incentivizing_CER_3-30 -11.pdf.

Helfand, M., Tunis, S., Whitlock, E. P., Pauker, S. G., Basu, A., Chilingerian, J., … Shepard, D. S. (2011). A CTSA agenda to advance methods for comparative effectiveness research. *Clinical and Translational Science*, *4*(3), 188–198.

Herdman, R. C., & Jensen, J. E. (1997). The OTA story: The agency perspective. *Technological Forecasting and Social Change*, *54*(2), 131–143.

Hlatky, M. A., Douglas, P. S., Cook, N. L., Wells, B., Benjamin, E. J., Dickersin, K., … Peterson, E. D. (2012). Future directions for cardiovascular disease comparative effectiveness research: report of a workshop sponsored by the National Heart, Lung, and Blood Institute. *Journal of the American College of Cardiology*, *60*(7), 569–580.

Hodgetts, K., Elshaug, A. G., & Hiller, J. E. (2012). What counts and how to count it: Physicians' constructions of evidence in a disinvestment context. *Social Science & Medicine*, *75*(12), 2191–2199.

Hoffman, A., Montgomery, R., Aubry, W., & Tunis, S. R. (2010). How best to engage patients, doctors, and other stakeholders in designing comparative effectiveness studies. *Health Affairs*, *29*(10), 1834–1841.

Hunink, M. M. (2005). Decision making in the face of uncertainty and resource constraints: examples from trauma imaging. *Radiology*, *235*(2), 375–383.

Husereau, D. (2010). Sentence first, verdict afterwards: using value of information analysis to inform decisions about pharmacogenomic test adoption and research. *Current Pharmacogenomics & Personalized Medicine, 8*, 167–170.

IOM [Institute of Medicine]. (2009). *Initial National Priorities for Comparative Effectiveness Research*. Washington, DC. Retrieved from http://nationalacademies.org/hmd/Reports/2009/ComparativeEffectivenessResearchPriorities.aspx

Johnson, A. P., Sikich, N. J., Evans, G., Evans, W., Giacomini, M., Glendining, M., … Perera, C. (2009). Health technology assessment: a comprehensive framework for evidence-based recommendations in Ontario. *International Journal of Technology Assessment in Health Care*, *25*(02), 141–150.

JOP. (2008). The Southwest Oncology Group: New Processes to Coordinate Increasingly Complex Clinical Trials. *Journal of Oncology Practice*, *4*(2), 78–80.

Kamerow, D. (2011). PCORI: odd name, important job, potential trouble. *BMJ*, *342*, d2635.

Keating, P., & Cambrosio, A. (2011). *Cancer on trial: oncology as a new style of practice*. Chicago: The University of Chicago Press.

Kim, S. Y., & Solomon, D. H. (2011). Use of administrative claims data for comparative effectiveness research of rheumatoid arthritis treatments. *Arthritis Research & Therapy*, 13(5), 129.

Kreis, J., & Schmidt, H. (2012). Public engagement in health technology assessment and coverage decisions: a study of experiences in France, Germany, and the United Kingdom. *Journal of Health Politics, Policy and Law*, 38(1), 89-122.

Larson, E. B. (2010). N-of-1 trials: a new future? *Journal of General Internal Medicine*, *25*(9), 891–892.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Lau, E. C., Mowat, F. S., Kelsh, M. A., Legg, J. C., Engel-Nitz, N. M., Watson, H. N., … Whyte, J. (2011). Use of electronic medical records (EMR) for oncology outcomes research: assessing the comparability of EMR information to patient registry and health claims data. *Clinical Epidemiology*, *3*(1), 259–272.

Le Pen, C. (2009). Is there a "continental" view of health economics evaluation? *The European Journal of Health Economics*, *10*(2), 121–123.

Lehoux, P. (2006). *The problem of health technology: policy implications for modern health care systems*. New York: Routledge.

Lehoux, P., Denis, J.-L., Tailliez, S., & Hivon, M. (2005). Dissemination of health technology assessments: identifying the visions guiding an evolving policy innovation in Canada. *Journal of Health Politics, Policy and Law*, *30*(4), 603–642.

Levin, L., Goeree, R., Sikich, N., Jorgensen, B., Brouwers, M. C., Easty, T., & Zahn, C. (2007). Establishing a comprehensive continuum from an evidentiary base to policy development for health technologies: the Ontario experience. *International Journal of Technology Assessment in Health Care*, *23*(3), 299–309.

Luce, B., & Cohen, R. S. (2009). Health technology assessment in the United States. *International Journal of Technology Assessment in Health Care*, *25*(S1), 33–41.

Luce, B. R., Drummond, M., Jönsson, B., Neumann, P. J., Schwartz, J. S., Siebert, U. W. E., & Sullivan, S. D. (2010). EBM, HTA, and CER: clearing the confusion. *Milbank Quarterly*, *88*(2), 256–276.

Luce, B. R., Kramer, J. M., Goodman, S. N., Connor, J. T., Tunis, S., Whicher, D., & Schwartz, J. S. (2009). Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. *Annals of Internal Medicine*, *151*(3), 206–209.

Manchikanti, L., Derby, R., Wolfer, L., Singh, V., Datta, S., & Hirsch, J. A. (2009). Evidence-based medicine, systematic reviews, and guidelines in interventional pain management: Part 7: systematic reviews and meta-analyses of diagnostic accuracy studies. *Pain Physician*, *12*(6), 929–963.

Manchikanti, L., Falco, F. J. E., Benyamin, R. M., Helm, S., Parr, A. T., & Hirsch, J. A. (2011). The impact of comparative effectiveness research on interventional pain management: evolution from Medicare Modernization Act to Patient Protection and Affordable Care Act and the Patient-Centered Outcomes Research Institute. *Pain Physician*, *14*(3), E249-282.

Markussen, R., & Olesen, F. (2007). Rhetorical authority in STS: Reflections on a study of IT implementation at a hospital ward. *Science as Culture*, *16*(3), 267–279.

Medicare Prescription Drug, Improvement, & Modernization Act (2003). §1013.

Merlin, T., Farah, C., Schubert, C., Mitchell, A., Hiller, J. E., & Ryan, P. (2013). Assessing Personalized Medicines in Australia A National Framework for Reviewing Codependent Technologies. *Medical Decision Making*, *33*(3), 333–342.

Moreira, T. (2007). Entangled evidence: knowledge making in systematic reviews in healthcare. *Sociology of Health & Illness*, *29*(2), 180–197.

------. (2012). *The Transformation of Contemporary Health Care: The Market, the Laboratory, and the Forum*. New York: Routledge.

Myers, E., Sanders, G. D., Ravi, D., Matchar, D., Havrilesky, L., Samsa, G., … Gray, R. (2011). *Evaluating the potential use of modeling and value-of-information analysis for future research prioritization within the evidence-based practice center program.* Methods Future Research Needs Reports, No. 5. Rockville: Agency for Healthcare Research and Quality (US). Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK62134/

Nass, S. J., Moses, H. L., & Mendelsohn, J. (Eds.). (2010). *A national cancer clinical trials system for the 21st century: reinvigorating the NCI Cooperative Group Program*. Washington, DC: National Academies Press.

NCI [National Cancer Institute]. (2009). NCI guidelines for ARRA research and research infrastructure grand opportunities: comparative effectiveness research in genomic and personalized medicine. Retrieved from http://www.cancer.gov/PublishedContent/Files/PDF/recovery/004 _cer_personalized_medicine.pdf

Neumann, P. J. (2012). What We Talk about When We Talk about Health Care Costs. New England Journal of Medicine, 366(7), 585–586. https://doi.org/10.1056/NEJMp1200390

NIH [National Institutes of Health]. (2009). American Recovery and Reinvestment Act. Supporting Comparative Effectiveness Research. Retrieved from http://www. cancer.gov/PublishedContent/Files/PDF/recovery/004 _cer_personalized_medicine.pdf

O'Rourke, K. (2007). An historical perspective on meta-analysis: dealing quantitatively with varying study results. *Journal of the Royal Society of Medicine*, *100*(12), 579–582.

OTA [Office of Technology Assessment]. (1976). Development of Medical Technology: Opportunities for Assessment. Washington, DC: US Government Printing Office

Patient Protection and Affordable Care Act (PPACA). (2010). §6301. (2010).

PCORI [Patient-Centered Outcomes Research Institute]. (2011). Rationale: working definition of patient-centered outcomes research. Available at: http://www.pcori.org/images/PCOR_Rationale.pdf.

------. (2012). 2011 Annual Report. Available at: http://www. pcori.org/assets/AnnualReport.pdf

Purmonen, T. T., Pänkäläinen, E., Turunen, J. H. O., Asseburg, C., & Martikainen, J. A. (2011). Short-course adjuvant trastuzumab therapy in early stage breast cancer in Finland: cost-effectiveness and value of information analysis based on the 5-year follow-up results of the FinHer Trial. *Acta Oncologica*, *50*(3), 344–352.

Ramsey, S. D., Barlow, W. E., Gonzalez-Angulo, A. M., Tunis, S., Baker, L., Crowley, J., … Hortobagyi, G. N. (2013). Integrating comparative effectiveness design elements and endpoints into a phase III, randomized clinical trial (SWOG S1007) evaluating OncotypeDX-guided management for women with breast cancer involving lymph nodes. *Contemporary Clinical Trials*, *34*(1), 1–9.

Ramsey, S. D., Crowley, J. J., Baker, L. H., Barlow, W. E., Burke, W., Garrison, L. P., & Tunis, S. R. (2009, October 29). Center for Comparative Effectiveness Research in Cancer Genomics (CANCERGEN). Poster session presented at: the Inaugural Meeting of the Genomic Applications in Practice and Prevention Network (GAPPNetTM): A National Collaboration for Realizing the Promise of Genomics in Health Care and Disease Prevention. Ann Arbor, Michigan.

Ramsey, S. D., Veenstra, D. L., Tunis, S. R., Garrison, L. P., Crowley, J. J., & Baker, L. H. (2011). How Comparative Effectiveness Research Can Help To Advance 'Personalized Medicine' In Cancer Treatment. *Health Affairs, 30*(12), 2259–2268.

Ray, T. (2009, October 21). CancerGen's First Comparative Effectiveness Trial to Study Oncotype DX in Node-Positive Breast Cancer. *GenomeWeb*. Retrieved from https://www.genomeweb.com/dxpgx/cancergens-first-comparative-effectiveness-trial-study-oncotype-dx-node-positive

Retèl, V. P., Joore, M. A., Linn, S. C., Rutgers, E. J., & van Harten, W. H. (2012). Scenario drafting to anticipate future developments in technology assessment. *BMC Research Notes*, *5*, 442.

Roland, M., & Torgerson, D. J. (1998). What are pragmatic trials?, What are pragmatic trials? *BMJ, 316*(7127), 285–285.

Schneeweiss S. (2007). Developments in Post−marketing Comparative Effectiveness Research. *Clinical Pharmacology & Therapeutics*, *82*(2), 143–156.

Scoggins, J. F., & Ramsey, S. D. (2010). A national cancer clinical trials system for the 21st century: reinvigorating the NCI Cooperative Group Program. *Journal of the National Cancer Institute*, *102*(17), 1371–1371.

Signorovitch, J. E., Wu, E. Q., Yu, A. P., Gerrits, C. M., Kantor, E., Bao, Y., … Mulani, P. M. (2010). Comparative Effectiveness Without Head-to-Head Trials. *PharmacoEconomics*, *28*(10), 935–945.

Smith, S. R. (2007). Preface. *Medical Care*, *45*(10), S1.

Sullivan, S. D., Watkins, J., Sweet, B., & Ramsey, S. D. (2009). Health Technology Assessment in Health-Care Decisions in the United States. *Value in Health*, *12*(s2), S39–S44.

SWOG. (2012). History. Accessed 22 Nov 2012. Retrieved November 22, 2012, from http://swog.org/visitors/history.asp

Teutsch, S. M., Berger, M. L., & Weinstein, M. C. (2005). Comparative Effectiveness: Asking The Right Questions, Choosing The Right Method. *Health Affairs*, *24*(1), 128–132.

Thariani, R., Veenstra, D. L., Carlson, J. J., Garrison, L. P., & Ramsey, S. (2012). Paying for personalized care: Cancer biomarkers and comparative effectiveness. *Molecular Oncology*, *6*(2), 260–266.

Thariani, R., Wong, W., Carlson, J. J., Garrison, L., Ramsey, S., Deverka, P. A., … Baker, L. H. (2012). Prioritization in comparative effectiveness research: the CANCERGEN experience in cancer genomics. *Medical Care*, *50*(5), 388-93

Trosman, J. R., Van Bebber, S. L., & Phillips, K. A. (2011). Health Technology Assessment and Private Payers' Coverage of Personalized Medicine. *Journal of Oncology Practice*, *7*(3S), 18s-24s.

Tunis, S. R., Benner, J., & McClellan, M. (2010). Comparative effectiveness research: Policy context, methods development and research infrastructure. *Statistics in Medicine*, *29*(19), 1963–1976.

Tunis, S. R., & Turkelson, C. (2012). Using Health Technology Assessment to Identify Gaps in Evidence and Inform Study Design for Comparative Effectiveness Research. *Journal of Clinical Oncology*, *30*(34), 4256–4261.

University of Michigan. (2010, April 12). U-M gets $63 million NIH grant for SWOG cancer trials network [press release]. Retrieved from http://www.cancer.med.umich.edu/news/swog-grant10.shtml. Accessed 31 July 2012.

Viswanathan, M., Ansari, M. T., Berkman, N. D., Chang, S., Hartling, L., McPheeters, M., … Treadwell, J. R. (2012). Chapter 5. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. *In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews* (pp. 112-29). Rockville (MD): Agency for Healthcare Research and Quality.

Whitlock EP, Lopez SA, Chang S, et al. (2009). Identifying, Selecting, and Refining Topics. *In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews* (pp.15-31). Rockville: Agency for Healthcare Research and Quality.

Wilensky, G. R. (2004). Implementing the MMA. *Healthcare Financial Management*, *58*(6), 30–32.

Wilson, E., & Abrams, K. (2011). From evidence-based economics to economics-based evidence: using

systematic review to inform the design of future research. In I. Shemilt, M. Mugford, L. Vale, K. Marsh, & C. Donaldson (Eds.), *Evidence-based Decisions and Economics: Health Care, Social Welfare, Education and Criminal Justice* (pp. 146–161). Oxford: Wiley-Blackwell.

Wong, W. B., Ramsey, S. D., Barlow, W. E., Garrison, L. P., & Veenstra, D. L. (2012). The value of comparative effectiveness research: projected return on investment of the RxPONDER trial (SWOG S1007). *Contemporary Clinical Trials*, *33*(6), 1117–1123.

WHO [World Health Organization]. (2000). A quick reference compendium of selected key terms used in the world health report 2000. Geneva: WHO. http://who.int/ health-systems-performance/docs/whr_2000_glossary. doc . Accessed 21 July 2008 (quoted in Einsiedel 2009, p. 187).

Interlude 1

In Chapter 3, we have characterized the field of comparative effectiveness research in an historically, thematically, and empirically-grounded analysis of primary and secondary literatures. As we saw in that chapter, CER has close links to other modes of healthcare evaluative research such as health technology assessment but also fundamentally differs in how it moves evaluation further upstream in the production and synthesis of biomedical knowledge. In concluding with a discussion and high level overview of CANCERGEN as a paradigmatic example of comparative effectiveness research in action, we were introduced to two specific movements upstream.

The first was in its participatory exercises of stakeholder engagement, which sought to move participation itself upstream in the process of prioritizing and designing CER studies in cancer care and prevention. The second upstream movement was the deployment of tools of economic evaluation which, similar to stakeholder engagement, were also applied to much earlier stage of assessing value -- here, the value of CER studies themselves.

With regard to the former, we suggested that the concept of a 'stakeholder' is in fact quite different from that of the 'public' and so we argued that the two terms should not be used interchangeably. However, elaboration of this argument was beyond the purview of that paper, while at the same time it was made at an earlier phase of writing and so the broader implications of this difference was still to be fleshed out. Over several iterations of grounded theory-informed analysis, this would ultimately become a more fully developed theory of *pre-public platforms* which I introduce in Chapter 4, and it is to this discussion that we now turn our attention.

Chapter 4

When participation goes according to plans: From public engagement to *pre-public* platforms

Abstract

The so-called participatory turn in sociological and STS scholarship has amounted to a tradition of in-depth studies of participatory mechanisms across scientific and biomedical policy- and decision-making. This corpus of work has succeeded at conceptualizing notions and qualifications of 'public' forms of action, while elaborating upon the many instances and consequences wherein 'publics' have failed to be enacted in the course of rectifying scientific controversies and uncertainties or in devising more democratic decision-making processes. Far less attention, however, has been paid to characterizing ways of engaging that occur beyond public legibility. The present paper seeks to begin filling this lacuna, drawing on ethnographic fieldwork with a project which sought to leverage a multi-stakeholder engagement process in streamlining the prioritization and design of biomedical research studies. Here, I deploy insights from pragmatic sociology -- in particular Thévenot's notion of *planning* -- to elaborate upon the notion of *pre-public* platforms. These are instances of decision-making typified by functional vocabularies, agencies, information formats, and a broader logic of confinement that resists escalating to moments of public justification and critique -- but where actors may nevertheless reflexively acknowledge a public situation on the horizon. As such, I argue that if analyses are to faithfully unpack the politics of participation in contemporary society, they are well served to consider a wider *ecology of engagements and (e)valuations* and the temporal nature of transitions between public situations and those which precede them.

1. Introduction

The proliferation of 'technologies of participation' has become an increasingly important topic in the field of science and technology studies as analysts set their sights on unpacking how various participatory mechanisms inform, or otherwise fail to impact, technoscientific and biomedical policy and decision-making (Soneryd 2015; Lezaun et al. 2017). This 'participatory turn' -- or what some claim is more like a participatory *re*turn -- within STS can be understood in the context of growing

critiques of technocratic modes of governance (Wynne 2007). In its midst, we find a profusion of collective problem solving approaches that appear both more attuned to broader societal concerns and better equipped to account for potential sociotechnical and ethical overflows that accompany the deployment of novel technological artifacts and policy interventions (Beck 1992; Bora and Hausendorf 2006; Jasanoff 2012). In some cases, these mechanisms emerge spontaneously around specific issue-areas and subsequently dissolve, only to reemerge later around new sets of issues; such is the case with social and protest movements (e.g. Epstein 2016). In other instances, we find varying forms of 'material participation,' which are constituted through individuals' engagement with the myriad objects and spaces that proliferate their daily lives and routines (e.g. Centemeri 2017; Marres 2012).

Still other technologies of participation, more 'procedural' than the former two, tend to be institutionalized as components of larger decision-making bodies (Degeling et al. 2015; Espeland 2000; Marres 2007). Citizens' conferences are a primary example of this now seemingly *de rigueur* practice technoscientific governance, with the so-called 'Danish model' of consensus conferences occupying something of an ideal-typical status here (Horst & Irwin 2009). These consultations typically occur in advance of or in tandem with policymaking and investment decisions around particular scientific or biomedical interventions and have been assembled to address an array of topics ranging from genetically modified foods, to the use of embryonic stem cells in biomedical research, to the regulation of nanotechnology (Bogner 2012; Harrison and Mort 1998; Irwin 2001; Laurent 2017). Gathering together policy and systems architects, scientists, and technical experts, as well as various *concerned groups* who stand to be impacted by the phenomena under consideration, these exercises aim to provide a forum for debate, deliberation, and shared learning around the problematized issue.

Be they *publics*, *citizens*, or *laypersons,* the act of labeling these latter concerned groups as such generally denotes their being made up of non-expert members of a (real or imagined) political constituency who assemble with some public issue in mind. In many cases, these groups are *invited* to participate, setting in motion a process of qualification whereby the inviters must make an evaluative judgment on who counts as a member of the public, as a citizen, or as a layperson. And yet within the body of STS scholarship on participation practices, one finds a tendency towards a vernacular if

not unreflexively normative usage of such labels. It is my contention that this leads to an overemphasis on the identities and ontological status of these groups, which in turn become a key concerns in the analyst's own evaluations of whether the exercise was properly executed given the qualifications stipulated at the outset and the ends in view. This comes, however, at the expense of a wider consideration of the nature of the situations in which these invited actors are participating, oftentimes precluding the possibility of problematizing the very notion of *public* forms of participation and thus of considering that a given participatory format may somehow be *other-than-public*. The present paper examines a still further category of participant frequently encountered in participatory settings -- that of *stakeholders* -- as a means of pushing further the analytic boundaries of STS scholarship on participation practices.

Empirically, my arguments draw from an ethnographic analysis of the Center for Comparative Effectiveness Research in Cancer Genomics (henceforth: CANCERGEN). This project sought to develop and refine a platform for prioritizing diagnostic technologies and designing protocols for clinical trials in precision cancer medicine, which would ultimately be conducted through the publicly funded cancer clinical trials research infrastructure in the US. A hallmark feature of this novel approach was a *stakeholder engagement* mechanism that incorporated the perspectives of a diverse group of healthcare constituencies into the to prioritization and design of large-scale clinical research studies, including decisions about what trials to conduct in the first place, as well as about the types of clinical and psychosocial endpoints that should be included in the design of those trials (Deverka et al. 2012b:182-3). This was enacted in the establishment of the External Stakeholder Advisory Group (ESAG), whose members were recruited to stand in as representative voices of a given group of decision-makers: patient representatives; consumers advocates; healthcare insurers; practicing clinicians; policy-makers; regulators; and the pharmaceutical and diagnostics industries. The guiding expectations of CANCERGEN were that 1) such a process would constitute a more streamlined and efficient prioritization and design platform; 2) that the platform would produce trials whose outputs would be more useful than single-investigator initiated trials, and would thus better account for the diverse informational needs of these various stakeholder groups; and 3) that these would combine to improve the well being of cancer patients.

Conceptually, I approach the work of CANCERGEN using the analytical lens of Thévenot's (e.g. 2007) *regimes of engagement* architecture. Within this framework, *stakeholder* is posited as a formatting of agency specific to a *regime of engaging in plans*, where actors' engagements are oriented by the *evaluative good* of realizing a planned action. This can be distinguished from other regimes of engagement, such as the regime of *public justification* and the more intimate *familiar* regime, which have their own unique evaluative goods and formattings of agency. Here, I take seriously the qualification practices by which CANCERGEN project organizers determined who counted as a legitimate stakeholder in this exercise and, in turn, how these stakeholders thusly defined were said to be distinct from *the public*. Approaching stakeholder engagement in this way insists that participation practices be understood in terms of relations that obtain between actors and their materially equipped environments, from which they derive the capacity to act. These environments may instill trust in actors that this is or that is the proper way to act, but they can equally lead to moments of inquietude and doubt -- a recognition of what is sacrificed by sticking to the plan, and how things could very well be otherwise (Blokker & Brighenti 2011). From this, I conclude that we are best suited to thinking of participation as occurring within an *ecology of engagements and (e)valuations*, where the ebbs and flows of action may well escalate into *public* situations, but where other logics are equally possible, including the logic of confinement through which actors simply and reflexively stick to the *plan* -- even while recognizing the potential for publicity somewhere on the horizon.

The next section of this paper walks back to review the wider participation literatures in STS as a means of problematizing its general bias towards *public* modes of participation. I then move on in Section 3 to a more detailed explication of the regimes of engagement framework before situating the elements of CANCERGEN's work in Section 4 as an instance of *engaging in plans*. The discussion in Section 5 resituates CANCERGEN as an instance of what I call an *pre-public platform*. This notion points to a particular type of situation that exists *prior to* what might be qualified as a *public situation*, and which operates according to a logic of confinement (cf. Boltanski & Thévenot 2006; Marres 2012). This, even in spite of the fact that the many project organizers and stakeholder participants whom I encountered in the field nevertheless displayed a significant and reflexive awareness about the 'societal context' of their work (e.g. Joly & Kaufmann 2009). The final section then offers some concluding reflections on the analytic work that such a concept does insofar as pushing social studies

of participation to consider a wider *ecology of engagements and (e)valuations* as well as the various affordances and oppressions that these different modes of engaging imply.

2. Problematizing *the 'public'* in the participatory (re)turn

Writ large, the STS literature has characterized the (re)turn to participation as a shift away from a paradigm centered on 'reliable knowledge' to one that instead stresses enhanced inclusiveness, accountability, and transparency of decision-making processes -- that is, the production of what Gibbons and colleagues (1994) have labeled *socially robust* knowledge. For Stirling (2008), social robustness is 'a substantive matter of the degree to which technological commitments are congruent with, or authentically embody, societally deliberated, publicly reasoned values, knowledges, and meanings' (272). Such 'social learning' processes 'are oriented toward informing the substance of social commitments themselves' and 'citizens are engaged as subjects rather than as objects of discourse' (ibid.). This stands in contradistinction to what he terms 'instrumental notions of social intelligence' that are instead aimed at 'conditioning the modalities for [the] implementation or presentation' of such technological commitments (ibid.).

Nowotny (2003) characterizes social robustness as occurring further 'upstream' in the research process; as being more 'contextual' than expertized pluralism; and as being fundamentally *relational* -- less a product than a *process* whereby expert knowledge is subjected to ongoing testing, resistance, and modification. She goes on to highlight three central aspects of this phenomenon. First, society becomes an 'active partner' in knowledge production by considering expertise beyond disciplinary specialization to include 'specific mixes' of other kinds of experiences and knowledges as well (ibid.:155). Second, the validity of such robustness is the outcome of a repeated sequence of expanding, modifying, and testing of social knowledge. Finally, this testing occurs at sites both within the laboratory and in the world beyond its walls, where 'social, economic, cultural and political factors shape the products and processes resulting from scientific and technological innovation' (ibid.). In sum, social robustness emerges in a 'problem-generating and problem-solving environment' where scientists and other knowledge-makers are said to 'encounter real and imaginary "interlocutors"'; this includes 'arrays of competing "experts" and the organizations and institutions through which they bring their knowledge and experience to bear on decisions taken, but also variously jostling "publics"' (ibid.:155-6).

This is all said to occur 'within the public arena' -- or, as Nowotny and colleagues 'would prefer to call it, the *agora*' (2001:23). The authors deliberately use this term, taken from the ancient Greek meeting place, as a model for 'embrac[ing] the political arena and the market place, and to go beyond both' (Nowotny 2003:156). The concept of the *agora* is said to hold an especially important place in the world of contemporary healthcare and biomedicine (ibid). It is in this space that we find 'new kinds of more open regulation' which are 'preceded by elaborate negotiations, mediations, consultations and contestations' (ibid.). Under the banner of 'patient and public involvement' (PPI), some jurisdictions now regularly deploy ad hoc consultations of this type, while others have gone so far as to establish standing advisory panels that aim to inform the assessment and evaluation of healthcare interventions and service provision (Harrison and Mort 1998; Knaapen & Lehoux 2016; Moreira 2015). A frequently referenced instance of this latter participatory format is within the National Health Service in the United Kingdom, which 'has repeatedly turned to the creation of standing committee-style spaces as the key route for public influence on local healthcare organizations (Stewart 2016:36). Nowotny and colleagues (2001) conclude that these types of engagements are 'no longer the domain of a relatively closed bureaucratic-professional-legal world of regulation' but are rather indicative 'of broader cultural-political movements embodying antagonistic forms of interaction which have become part of the repertoire of how novel technologies are embedded and research products come to be accepted and used in wider social contexts' (23).

A closer look at this framing, however, reveals a particularly dichotomous conceptualization of what it means for something to be (a) *public*, which I argue poses certain limitations on social scientific analyses of participation practices. For instance, the *agora* is positioned as being coextensive with the *public arena* -- the implication here being that this is an altogether different locus than the more private experimental spaces of the laboratory or bureaucratic environments of policymaking. On the other hand, we find discussion of *publics* posited as real or imagined groupings of societal actors, characterized as being distinct from *scientists* and other types of *experts*. So despite the widespread influence of the social robustness model in science and technology studies, it nevertheless produces a twofold challenge for studying participation practices in technoscience, biomedicine, and beyond. The French sociologist Laurent Thévenot (2014a) has recently framed this problem as, on the one hand, stemming from an 'ambiguity of the double meaning of a "public," as

an *audience* and as the *public realm of a political community*' which leaves many observers and participants 'puzzled and uneasy due to the confusion between different involvements' (153, italics mine). On the other hand, he says, scholars of participatory democracy 'usually concentrate on access to a discursive public space of argumentation,' and yet close empirical investigation reveals 'other modes of expression' and of achieving commonality (ibid.)

Thévenot concludes that '[c]urrent analyses of public spaces fail to grasp this complex situation' and so 'we need an analytical framework that is capable of grasping practical engagements below the level of publicity, and ways in which human beings engage with the world all the way down to a level of close proximity' (ibid.). This need becomes all the more urgent when the analyst is faced with a participatory exercise that exhibits many of the qualities of what is described above as the production of socially robust knowledge -- occurring in a space resembling the *agora*, to which I would add Callon and colleagues' similar notion of the *hybrid forum* here, too -- but nevertheless notices a reflexive acknowledgment among the participating actors that they are explicitly *not* conducting a form of 'public engagement' in knowledge production (Callon et al. 2009; Rabeharisoa & Callon 2003).

Such was the case with CANCERGEN. Of particular relevance for the present analysis is the fact that CANCERGEN project organizers explicitly stated that they were *not* attempting to engage *the public*. Rather, in working to capture the varying perspectives of these decision-makers, they instead claimed to be engaging a different category of actor -- *stakeholders* -- which they operationalized as those with 'self-interest in a given issue' and which therefore implied that 'their involvement in a topic is seen as both rational and more likely to contribute to the quality and legitimacy of subsequent actions' (Deverka et al. 2012b:183, italics mine). The CANCERGEN initiative is thus an instructive case for exploring the fruitfulness of extending social scientific analyses of participation practices -- especially in ways that might better capture the diversity of modes through which individuals engage with each other and their material environments, and work to achieve commonality in formats that are somehow other-than-public yet not wholly reducible to ostensibly 'other' categories like 'bureaucratic' or 'private.' Accordingly, in the next section I turn to a more detailed treatment of the regimes of engagement framework such that we can better comprehend

Thévenots call for deploying such a novel conceptual apparatus, before deploying it in an analysis of CANCERGEN's work.

3. The public and its other(s)

3.1 Reframing publicity through a pragmatist lense

To speak of analyzing participation practices that go *beyond* the public requires an explication of what is meant by *public* in the first place. As pointed out in the introduction, the word has taken on a somewhat vernacular usage in the STS literature, referring to a myriad of different entities: a kind of *space*, an *audience*, or some other *grouping of concerned actors*. However, Thévenot's usage -- which I draw from in the present analysis -- has a more specific referent, which has grown out of the approach most commonly referred to as pragmatic sociology (e.g. Barthe et al. 2013; Benatouîl 1999; Blokker 2011; Dodier 1993; Guggenheim & Potthast 2011).

The approach to public issues in this body of scholarship is not limited to a focus on formal political institutions, per se. Rather, for these authors, 'it is because people have the ability and willingness to justify their viewpoint to others that they reach beyond the proximal domain of their particular situation' and so situations come to have a public quality 'because they entail a basic orientation towards "generality"' (Eulriet 2014:418). Work emanating from this research program thus fleshed out a particular view of the public that considers a heavy reliance on *conventional forms of equivalence*, where justification and critique are mobilized in such a way that requires a certain detachment from more personalized concerns and works to 'encapsulate the institutional power of bringing together by way of a common abstraction' (Thévenot 2011:37). The most detailed description of such an idea is found in Boltanski and Thévenot's (2006) *On Justification*, where the authors specify six legitimate *orders of worth* that actors routinely draw on when publicly calling out situations of injustice. They label these orders industrial (efficiency); civic (solidarity aiming at greater equality); domestic (trust and reputation based on custom and tradition); market (competition); inspiration (creative); and fame (public opinion). In their totality, these six orders can be approached as 'practical and ideological devices built to avert certain troubles and bad consequences' that may threaten communal life (Stavo-Debauge 2012:9). Together, they comprise a 'horizontal pluralism' (Eymard-Duvernay et al. 2005) of conventions that facilitates the highest order of public coordination and provides the tools for actors to objectively assess the worth of people and things.

Each order (*monde*, or 'world') on its own, however, is traced to a particular foundational work in political philosophy and accordingly signifies a specific configuration of the *common good* -- that is, a guiding orientation for how humans are to live together in a 'common humanity' (Thévenot 2002). Along these lines, Centemeri observes that

> the public form of action in Western society has been socio-historically conceived and operationalized through a variety of cognitive artifacts, disciplining devices, and technologies to support the construction of both the public sphere and the individual. According to this construction, in order to be publicly justifiable, a mode of valuation must rest on a universally legitimate underlying good, meaning that this good must potentially benefit humanity as a whole: a truly "common good"' (2017:101).

That is, each order implies an attendant set of qualifications that are deployed within public disputes (Centemeri 2015:304; Thévenot 2014b). Qualification, in this vocabulary, refers to the process by which actors engaged in critical situations operate a set of criteria upon (other) people and (material) objects. In doing so, these entities undergo a transformation as actors go about testing the relevance or fitness of these entities for a given situation (Boltanski & Thévenot 2006).

In this sense, qualification serves as the basis for evaluations 'in which coded forms and qualifications pretending to public legitimacy are grounded' and works to 'raise the generality' (*montée en généralité*) of arguments by framing phenomena present in a given situation in terms of this wider appeal to to a common good (Thévenot 2011:43). In doing so, actors 'make visible a link between a cognitive framing and an ideal of justice,' which enables them to 'cross micro and macro levels of organization' in their attempts to negotiate and resolve a conflict (Moreira 2012:37). By qualifying for a given form of coordination, the entities in turn derive a durability from this capacity to coordinate. But in this capacity is also contained a sacrifice, as qualifying for one particular mode of coordination oftentimes eclipses the possibility of qualifying for an alternative, competing mode (Thévenot 2011:41). In these public disputes, one world often 'wins out' over another, but it is also possible for *compromises* to occur. Such compromises are, however, costly and fragile arrangements which nevertheless enable a single entity to qualify according to multiple visions of the common good.

Pragmatic sociology thus posits an 'understanding of publicness as a specific quality of action and

of the modes of valuation that guide it,' pointing not just to a type of space, a mode of discourse, or a grouping of human actors, but an entire set of relations linking the cognitive, the moral, and the material in a complex nexus of evaluation and critique (Centemeri 2017:101). In doing so, it falls in line with the wider body of STS scholarship on participation which, in contrast to more discursive approaches found in political science and democratic theory, instead considers 'associations of material and social constituents' (Marres 2007:762). It also bears some similarities to to early 20th century pragmatist conceptualizations of the public, notably put forth by John Dewey, who defined the public as 'consist[ing] of all those who are affected by the indirect consequences of human action, to such an extent that it is deemed necessary to have those consequences systematically cared for' (quoted in Marres 2012:45). This line of thinking was developed in specific opposition to the views of Walter Lippman -- one of Dewey's contemporaries and key interlocutors -- who 'dissolved the public into a large stakeholder grouping, which is less interested in the issues at hand than stakeholders are' (Marres 2012:53). For Dewey, then, publics are 'not just another, larger, stakeholder community whose concerns must be taken into account' but are rather 'called into being by issues are *morphologically different from communities of stakeholders*, those that consist of issue experts and actors with a direct stake in the issue at hand' (ibid:58, italics mine).

Marres argues that this way of approaching publics transcends the grammars and practices through which publics are organized and refers 'to a particular *modality* of being implicated in inherently dynamic formations, which stand out first and foremost for the requirement of some kind of collective action upon them' (ibid.:45). Deweyian publics thus form up around 'problematic situations' that resist the problem solving strategies most familiar to actors, a type of trouble that Marres (drawing from Woolgar) calls *ontological trouble*. Their formation, moreover, entails an

> ostensible disruption of working routines, and the opening up of a space of articulation as a consequence. When actors experience harmful indirect effects, they are transformed from ordinary actors, caught up in habitual ways of doing things, into participants - or at the very least, 'implicants' in problematic assemblages (Marres 2012:48).

As Marres explains of this perspective, public involvement is understood as an opening of issues to *political outsiders*. Meanwhile, the *public* can be approached as a *method* that 'work[s] in favor of the coming of the community. The community gains [its] individuality and distinguishes itself through

the uniqueness of a will, in promoting common goods and shared values with which it insures its preservation and renewal' (Stavo-Debauge & Trom 2004 quoted in Stavo-Debauge 2012:7). Deweyian pragmatism, in sum, characterizes public affairs in terms of a 'resistance to institutional settlement' where one must 'distinguish between issue articulations that open an affair up for public involvement, and those that prevent this from happening. [...] As public issues depend on outside involvement for their settlement, the difference between "publicizing" and "de-publicizing" issue articulations is of central importance' (2007:772; cf. Stirling 2008).

Marres' rendering of Dewey's notion of the public can, to an extent, be read alongside Boltanski and Théveot's own conceptualization in that the former, in its focus on the formation of public issues, is perhaps better equipped to describe how publics might emerge in this process. The latter, in turn, provides a set of tools for analyzing the enactment of public situations, the socio-material dispositifs they imply, and the various grammars that actors draw on when justifying their arguments for one or another vision of a good community (Holden et al. 2013; Stavo-Debauge 2012). However, there are certain marked differences between these approaches as well. Primary among them is the fact that that Dewey's vision of the public -- understood as the Great Community -- appears normatively tied to a single common good, while the very starting point of Boltanski and Thévenot's approach is that there is great variability between the public goods that are at stake in public controversies (Holden et al. 2013; Thévenot 2011).

A second source of normativity is found in Marres' own explicit 'normative appreciation of material publics' (2012:54). This contributes to a benchmarking for whether a situation is *really public*, with publicity contrasted to some *other*, be it institutional habits or bureaucratic decision-making (e.g. Espeland 2000). Analytically, the emphasis that Dewey (and, by extension, Marres) places on the role of 'political outsiders' in turn tends to situate the analyst as an adjudicator whose role is to measure the extent to which outsiders' perspectives have been adequately accounted for in public situations and whether a public has been adequately produced. This comes at the expense of closer scrutiny of how the tests by which actors evaluate the problematic situations in which they find themselves are constructed, while sidelining analyses of the qualification processes through which actors gain their capacities to act in these situations.

But the most important limitation for the purposes of the present analysis is that Dewey's *affected*

groups and Lippman's *interested stakeholder groups* are positioned as competing visions of the public, the former politicizing and the latter de-politicizing. In this way, it is not only that Dewey's public lacks a consideration of 'the confrontation of a plurality of the most legitimate species of the common good,' but it also does not consider the possibility of other ways of engaging that may not be immediately prepared to compose a 'shared common good' -- of which Lippman's stakeholder model serves as a key example (Stavo-Debauge 2012:18). To account for this variability absent in Deweyian notions of the public, Thévenot has in more recent work gone on to reappraise his work on public justification discussed above. Motivating this move is an understanding that, although orders of worth 'may account for the relation between cognitive and political demands,' evaluations according to the common good do not 'exhaust inquiry into the cognitive and evaluative formats used in social life' (Thévenot 2007:415).

3.2 Regimes of engagement and planned action

If we wish to more fully capture the variety of such formats that exist in contemporary societies, one must go beyond those conventional forms to consider the 'more dynamic notion of *regime[s] of engagement*' which 'emphasises that the human capacity at stake depends on the disposition of the material environment as well as of the person,' where individuals derive their confidence not only from their own commitments to a given situation but also to a 'dependence on a properly disposed environment' (Thévenot 2011:36/48, italics mine). Engagement, understood as a 'quest for certainty' sustained through one's relationship to their material environment, specifies the conditions under which one gains confidence in her capacities for engaging with herself and with the extant world in pursuit of a particular good. Each regime specifies a *good* that orients action; a specific designation of *agency*; a formatting of the *environment*, which is itself populated by a corresponding arrangement of *objects* and *information formats*; and a corresponding temporal horizon.

Within this conceptual schema, Thévenot resituates the approach he developed together with Boltanski the as the *regime of public justification*. Here, orders of worth predominate within such public engagements, informing the evaluative good orienting action by way of people and objects qualifying for *worth*; the agency of actors, who act in the name of the common good so specified; the conventional formatting of information and objects, as well as of the environment. This regime demands of actors that they express themselves using a *grammar of plural orders of worth* that is publicly

recognizable; arguments must be 'formatted in relation to a characterization of the common good,' which also applies to the forms of supporting 'material evidence' such that they can be evaluated according to one or the other competing orders of worth, which have varied temporal horizons (Thévenot 2014b:16). The regime of *familiar engagement* is perhaps most distal to what might be called the 'public sphere.' Turned to the past, the good particular to this regime is ease of accommodation with one's surroundings while agency is understood in terms of the 'personality with attachments' (Centemeri 2015:305). This regime assumes a *grammar of personal affinities to multiple common-places* which recognizes both individuals' personalized and idiosyncratic attachments to specific local spaces while at the same time enabling a 'broader commonality' to be formed within them by the actors who converge there (Thévenot 2014b:10).

The *regime of engaging in a plan* (or the regime of planned action) is still different from these first two regimes, and occupies something of an intermediate position between them. The evaluative good guiding action and coordination here is the achievement of a stated objective such that reality is

> grasped with respect to successful realization of the plan, which implies that it takes shape as a function instrumentally appropriate to the plan of action. […] The plan intention cannot be experienced without recognition that environmental components have a functional capacity -- this is what ensures the type of guarantee particular to this regime (Thévenot 2007:417).

Actors' agency is understood in terms of the willing and autonomous subject who derives her capacity to act from a functionally supported environment, as a means for projecting oneself into the future. Importantly, this agential framing finds its 'ideal type' in the form of the so-called *stakeholder*, who is expected to 'be "practical," "realistic," and "effective" against what is "absolute or the ideal"' (Thévenot 2015:93). Discourse is populated by a vocabulary of 'necessities' and 'utility,' and environments are correspondingly equipped: functional tools, objects, and information formats all serve to enable actors to communicate about and evaluate the planning and execution of a given course of action without requiring them to engage on the basis of conventional qualifications that situations of public dispute otherwise demand (Thévenot 2011:54). The grammar suggested by this regime is what Thévenot refers to as the *liberal grammar*, which 'tames disagreements and channels them within the boundaries of interests or preferences expressed by individual choice' (Thévenot 2014b:10). Pattaroni (2015) has further pointed to how this grammar differs from regimes of public

justification and familiar engagements, implying instead a form of 'contractual commonality' and a 'process of delimitation of the contractual objectives and the negotiation of interests' which enables actors to 'hierarchize' a set of 'established objectives' (166). It is not actors who 'hybridize' in instances of contractual commonality, in other words, but rather that the planned engagement 'produces hybridization,' thus 'allow[ing] for coordination and interaction between […] agents without them being asked to abandon their respective identities, skills, or projects' (Muniesa & Callon 2007:174)

Having mapped these three regimes and their corresponding grammars in short form, a certain notion of 'publicity' (*publicité*) comes to the fore alongside a clearer understanding of the diversity of situations in which humans and their environments together produce the preconditions for action. Each regime can be approached in terms of the particular good in play; the qualification of actors' agency; the formatting of information; and the equipment of the environment. These elements, together with the implied grammar of each regime, also enable the analyst to thus differentiate between situations of action that display a correspondence to a given mode of engaging with the world. Action, from this perspective, is understood as dynamic and pragmatic, and presupposes an ability of actors to shift between regimes in light of the demands that different settings might impose. In this regard, borrowing from Thévenot's toolbox leaves us well placed to better distinguish situational characteristics that might resist those most conventional 'public' qualifications. The advantage of the regimes of engagement approach becomes especially clear when the analyst is faced with situations that in many ways reflect instances of 'public involvement' discussed in the preceding sections of this paper, but that exhibit peculiar characteristics pointing in some other direction. Along these lines, it is my contention that the stakeholder engagement practices deployed within the CANCERGEN prioritization process stand as an instance of *engaging in planned action*, rather than of either public justification or familiar engagement, a discussion we now turn our attention to.

4. Prioritizing plans in the CANCERGEN platform

4.1 Choosing a trial: the evaluative good

The Center for Comparative Effectiveness Research in Cancer Genomics was a three-year initiative, first funded by the US National Cancer Institute in 2009. A multi-institution collaboration, the program sought to leverage a diverse structure to execute 'a consensus process with multiple

stakeholders [to] develop a comprehensive evaluation, assessment and design process to prioritize emerging cancer genomics technologies' for conducting clinical trials (Deverka et al. 2012a:361). The endeavor was quite a complex undertaking in terms of both its organizational makeup and the diversity of deliverables the project pursued. It involved the collaboration of project members from four different institutions, including an academic cancer hospital; a health economics and outcomes research department at a large state university; a nonprofit organization specializing in stakeholder engagement in healthcare; and SWOG, a nationwide infrastructure through which publicly-funded cancer clinical trials are run. In addition to these four organizations, an additional element of the project was the External Stakeholder Advisory Group (ESAG), which was comprised of a group of individuals who were invited to participate as representatives of various decision-making constituencies in the American healthcare system.

Although the project was largely framed as developing a *process*, demonstrating the value of that process -- as an improvement over other ways of prioritizing and designing clinical studies -- rested on meeting a series of milestones over the course of the project. The first was for the stakeholders to come to a consensus on a single diagnostic technology on which to conduct a large-scale clinical trial. Once this decision was made, the technology would then carry forward to a subsequent participatory design process in which the stakeholders would provide input on the various clinical and socio-emotional endpoints to be measured in the trial. Then, in a final stage, that trial design would be handed off to SWOG, who would use its national network of affiliated investigators to evaluate the clinical usefulness of that test. This envisioned cascade of milestones within CANCERGEN sensitizes us to the overall orientation of the project -- what, along with Thévenot, we may call the *evaluative good* of accomplishing a stated plan of action, and which corresponds to a specific regime of *engaging in plans*. Along similar lines, the mutual engagement takes the shape of a *joint project*, where project organizers together with the stakeholders share a commitment to the prioritization and design exercises with a specific end in view. But as we have seen in the previous section, regimes are not a matter of evaluative goods and mutual engagements alone but are rather situational *dispositifs* that also stipulate formattings of agency, information, objects, and material environments which in turn feed back onto a given engagement as a means of instilling trust that the particular way of engaging is appropriate or convenient for the purposes at hand.

4.2 Qualifying (for) engagement: defining, recruiting, and selecting

Project organizers largely framed the value of CANCERGEN's work -- prospectively as well as retrospectively -- as being rooted in a novel approach to prioritizing and designing clinical trials of diagnostic technologies. As we have seen above, front and center in this process was a *stakeholder engagement* component. By convening the External Stakeholder Advisory Group, it was expected that 'the quality of research will be enhanced, research will produce evidence better aligned with healthcare decision-making needs, and study results will be more likely to be translated into clinical practice' (Deverka et al. 2012a:359). But much of the early work in CANCERGEN was devoted to operationalizing what *stakeholder engagement* would mean in their specific local enactment. One of my interview respondents, a project organizer, noted the challenges of defining stakeholder involvement. In his view, it was 'hard to draw the line' as to 'when is something a formal stakeholder engagement' and thus operationalizing it isn't 'really black and white [...] and that's the problem.' But as project organizers worked to enact this thing called stakeholder engagement, they positioned their own approach relative to other known participation practices and methods in healthcare.

The aforementioned project organizer also stressed during the same interview that he didn't 'think stakeholder [engagement] in research […] is brand new. It's just been *formalized* and given a name, so it'll be interesting to see how it evolves and interesting to see some comparisons between where this happens elsewhere and what's happening now.' He went on to enumerate several possible models for stakeholder engagement:

> One is, there's a whole field of community based participatory research (CBPR), [...] the concept that people who are subjects of studies and involved in the research themselves, they should have some input into what gets studied, and some say about study designs, and so it's much more a *collaborative* type arrangement. There's also another model that's interesting to think about, in private industry and [the] pharmaceutical industry. They do this all the time -- this is their bread and butter […] They use advisory groups more as *marketing tools* but they do a lot of advisory group stuff during early stages of development to find out, 'OK, well, if we designed a trial like this and these are the outcomes, would this change clinical practice or not?' They invest -- and it's market research.

Comparisons between the CANCERGEN stakeholder engagement process and community based participatory research initiatives in the US are also drawn in an early paper co-authored by several project organizers which, along with an additional reference to similar 'public involvement initiatives' in the UK's National Health Service, note the importance of these different practices in their 'recognizing and fostering the role and unique contributions of patients and communities to participate in the research process' (Deverka et al. 2012a:359).

From these excerpted passages emerges a certain interpretive flexibility (e.g. Bijker 1995) around the notion of stakeholder engagement, both within CANCERGEN as well as in the wider participatory landscape in healthcare. Project organizers highlight this flexibility in a second published manuscript that includes a review of the literature on participation methods conducted at the project's outset, and which was used to inform their framework for stakeholder recruitment (Deverka et al. 2012b). Here, again, the authors note persistent conceptual difficulties surrounding the notion of 'stakeholders' within the wider published literature: they range from 'groups with expert knowledge' to 'patients or primary caregivers' to 'individuals who speak for patients/consumer,' and even 'references to stakeholders with personal experience with a health condition often involve inconsistent use of terminology, including "patient", "consumer", and "user"' (Deverka et al. 2012b:183).

Implied in this flexibility is the necessity of judging who counts as a legitimate stakeholder in a given participatory exercise and thus who is granted access to the decision-making table. Would the category of *stakeholder* include cancer patients entering a clinical trial, who stand to more immediately benefit from direct involvement in co-determining the study's objectives? Or would they be more along the lines of members of an advisory group convened by the marketing department of a private pharmaceutical or diagnostics firm, recruited to facilitate the firm's strategy to 'singularize' its product (Callon & Muniesa 2007:1233, cf. Lezaun 2007)?

Project organizers were in fact quite quite clear about how they defined a stakeholder for the purposes of their project:

> Although the categories can appear ambiguous and overlapping, the critical distinction for our purposes is that the terms 'public' or 'citizen' are best reserved for individuals without a direct interest in a particular healthcare issue. *Stakeholders have been distinguished from 'the public' because they have self-interest*

> *in a given issue*; therefore, their involvement in a topic is seen as both rational and more likely to contribute to the quality and legitimacy of subsequent actions (Deverka et al. 2012b:183, italics mine).

This notion of the self-interested representative thus came to occupy a central role in how stakeholders were recruited and engaged throughout the duration of the project, a point openly and frequently referred to in its published material as well as in the everyday engagement practices. And yet 'recognizing that each project will be unique in its specific stakeholder perspective requirements,' project organizers outlined a limited number of discrete categories of actor which were deemed the most relevant interested parties for the purposes of recruitment and selection of individuals for External Stakeholder Advisory Group: (*a*) patients and consumers; (*b*) healthcare providers; (*c*) industry; (*d*) purchasers and payers; and (*e*) policymakers and regulators (ibid.:184). More than simply being a member of one of these constituencies, however, there was an additional set of qualifications which ESAG nominees were expected to meet: they were to be 'senior representatives of their respective organizations with a working knowledge of genetics, personalized medicine or oncology,' who also possessed 'excellent communication skills' and would be 'abl[e] to commit the requisite time for a 2-year study' (Deverka et al. 2012a:360). A still further 'desirable' criterion was for nominees to have 'practical experience' in evaluating biomedical technologies, especially genomic tests (Thariani et al. 2012:390).

Given the specificity of these combined requirements, it should come as little surprise that interviews revealed that several of the stakeholders understood their recruitment not only in terms of whatever knowledge, expertise, experiences, and/or organizational affiliations they brought to the negotiations, but that their selection was also a function of some preexisting relationships they had with CANCERGEN project organizers. In other words, stakeholders' agency was constituted within a dialectic of representativeness-and-convenience. Consider, for instance, the following excerpt culled from a transcript of the first in-person ESAG meeting, held in mid-2010. The would be the first time that the stakeholders would have interactive discussions about possible options for prioritizing and designing a clinical trial on a diagnostic technology, and so began with a project organizer reviewing the expectations for the day's activities. As she detailed the deliberative methods

(more on this later) that would be used and the importance of having an ethical and balanced conversation, she reminded the stakeholders that

> as much as possible, we want the representation of your participation as *a representative of your stakeholder group*, *not you as an individual or your individual organization*. I certainly appreciate that [...] one-word stakeholder groups are not monolithic. We can't pretend to speak for all payers and that certainly would be a naïve idea, but as much as possible, if you could bring [this] perspective.

The representativeness of the stakeholders seemed to take on particular importance in the evaluative situations (Ehrenstein & Neyland 2016) during which they were asked to rank and vote on candidate technologies during the prioritization process. One instance of this came during the penultimate round of stakeholder voting at the June 2010 ESAG meeting. At this point, stakeholders were asked to choose only the three highest ranked diagnostic technologies from the shortlist of six with which they were initially presented; these three tests would then move on to the final round of voting, which was expected to result in consensus on a single technology to carry forward as the basis for designing a clinical trial to be conducted by SWOG. Here, the meeting facilitator emphasized that her

> next question is a serious question. And this is not necessarily easy because I know *some of you wear multiple hats*. But if you could identify which of the stakeholder groups you most consider yourself to be [...] The question about stakeholders is very important because it allows to see the results by stakeholder group [...] But I think *what is interesting is just to see if there are differences between stakeholder responses and investigator responses* because the real promise behind comparative effectiveness research [...] is *the hypothesis that the investigator-led research is not where we need it in terms of getting the types of things [informing] clinical studies* [...] So just as an overall, these three tests were the top rated tests [...] This is the result from the overall impression. You can see that *it's entirely clustered* with some clear winners. There were also some clear losers [...] But it's interesting too because clearly I think that because there is such clustering in here that some of you were considering some domains, like for example clinical validity in the test -- I'm just throwing that out as a possibility as more important than others. And so it's interesting and today we want to explore a little bit about your perceptions about the weighting and

your perceptions about which of these criteria are more important.

A still further discussion about stakeholder representation arose during the second in-person ESAG meeting, which I attended in March 2011. Here, I listened in on the following exchange between several of the stakeholders and project organizers as they addressed some of the difficulties and frustrations about how things were being communicated between the project's different organizations, as well as between the project organizers and the stakeholders:

> *[Consumer]*: it's always good to remember our role and to take a step back as stakeholders and remember that we don't really represent the groups. We represent one opinion. I mean I certainly don't represent all consumers, or even close [...]
> *[Payer]*: I think that is very true. In fact you have sitting on the payer side here two known outliers so I know there are colleagues of mine from other health plans that would disagree with what I say, probably some of them fairly strongly [...]
> *[Industry]*: That's right, that's why I was a little cautious this morning when asked who I represented. I guess I could potentially represent the entire therapeutics and diagnostics industries: a homogeneous, altruistic, forward-looking group of entities. But in fact, what you are getting is a mixture of me putting in perspectives that I think represent the kinds of issues that might come up from those perspectives, but tempered or contaminated with whatever your opinion is [of] my own personal views
> *[Unidentified]*: Exactly, which are a different value --
> *[Organizer]*: Well, that was a very helpful and good conversation, and we appreciate it. We understand that we didn't choose you to be a representative sample of the various stakeholder groups. It's a [purposive] sample, and what we are looking for is a *consensus*.

What we find here is a particular enactment of a 'stakeholder' category, qualified as a representative individual with direct material interest in the outcomes of the research they were recruited to participate in designing. Stakeholders, moreover, were conceived as a type of 'political outsider,' distinct from the category of *investigator* -- that is, the Principal Investigator of a research study who typically designs, submits for funding, and oversees the conduct of a given clinical trial.[11]

And yet, despite the seriousness with which CANCERGEN project organizers treated the representativeness of the various predefined stakeholder categories -- in the up front work of

planning the exercise and laying out the qualifications of who would count as legitimate stakeholders in the participation process -- it appears that, in their situated actions (Suchman 1987) both project organizers and the stakeholders themselves expressed a reflexive awareness that stakeholders' identities were not in any way fixed. In the exchange(s) excerpted above, it is evident that even within the course of a single day, and sometimes even in a matter of minutes, a single stakeholder could alternate between expressing multiple points of view. This was the case with the 'payer' stakeholder, who speaks of moving between representing the entire insurance industry and his own, perhaps more idiosyncratic, views that privileged quality of life concerns over his company's bottom line in his opinions about whether and how to incorporate certain endpoints into the final trial design; the same was true of the stakeholder invited to represent the 'industry' perspective who notes the 'mixture' of perspectives he had contributed to the deliberations. .

With or without what we might call this more pragmatically flexible treatment of representativeness, the agential formatting of CANCERGEN stakeholders was nevertheless tied to a predominating *liberal grammar* that emphasized choices and available options to be deliberated upon, with stakeholders communicating vis-a-vis expressions of their 'own personal views' and of 'representing one *opinion*' (even if a 'contaminated' one), while project organizers took great interest and care in the 'weighting' these different 'perceptions' as they went about facilitating the deliberations in order to further explore the drivers of these preferences broken down by stakeholder category. Different from and collectively more informative than the intellectual motivations of a single investigator who designs a study independent of such a wider engagement process, project organizers expected this panoply of traditionally unaccounted-for stakeholder perspectives to be key contributions in achieving one of the main objectives of the project: the design of a robust clinical trial, informed by the knowledge and expertise of patients, consumers, industry, payers, and regulators alike. It would also serve as a way to demonstrate the proof-of-principle of this novel process for prioritizing and designing cancer clinical trials, which could serve as confirmatory evidence in making the case for scaling up the process across the Cooperative Group Program, or even for other decision-making settings.

In sum, the twofold evaluative goods of reaching 'consensus' on a single diagnostic technology and trial design, and of demonstrating that the process was a workable one, were closely coupled

with the formatting of participants' agency as *stakeholders* and the deployment of a *liberal grammar* as a means of communicating. From the vantage point of the *regimes of engagement* architecture, these phenomena work to sensitize our analysis to a particular mode of coordination in play within CANCERGEN, one that is suggestive neither of public nor familiar engagements, as previously described, but rather a situation that can be qualified as *engaging in plans*.

4.3 Informed environments and objects (of contention)

There are still further elements of the planning regime's situational *dispositif* we have not yet fully accounted for heretofore in our analysis of CANCERGEN: namely, the *environment* in which the stakeholder deliberations took place as well as the *objects* and *information formats* which served as functional markers during these engagements. Recall the 'multiple hats' excerpt above, for example, which we have used to demonstrate the importance that project organizers placed on understanding and stratifying the perspectives of the different stakeholder groups vis-a-vis the different options for diagnostic technologies. This is but one example of project organizers' routine attempts to quantify how the different categories of stakeholders were ranking the technologies at these different moments of prioritization and voting over the course of the project. Results of these voting exercises, which were conducted online prior to the meetings as well as during the in-person meetings, were circulated among project organizers as well as shared back with the stakeholders themselves. In the latter case, each of the stakeholders was issued a remote clicker so that they could anonymously respond to questions about their perspectives and choices on the set of technologies, the results of which were displayed on a screen at the front of the room.

This panoply of online survey instruments, electronic voting systems, and remote clickers were key means of producing information over the course of CANCERGEN, while the sharing of survey results -- via PowerPoint presentations projected on a screen at the front of the room during ESAG meetings, for instance -- were key means of conveying that information. Distinct from the conventional markers of public engagement and the personalized markers of familiarity, we are quick to notice the *functional* formatting of these tools and information that proliferated the environment -- something 'as much needed as the intentional agency attributed to human beings' for facilitating coordination in carrying out a plan (Thévenot 2002:73). Referring to the functionality of these entities here is to mark the way they served to reinforce the overall objective of choosing a

technology to trial, while also feeding back on to stakeholders' capacity to act within this situation, where they were asked to effectively play the 'role' of a given stakeholder group (cf. Goffman 1956) and impelled to speak in a liberal grammar of choices and options -- to which these functional objects and information formats gave material form.

The introduction of an esoteric and hybrid form of economic-decision analytic modeling called value of information analysis (VOI) during the CANCERGEN prioritization and design process constitutes a second type of functional object and corresponding information format which aimed to facilitate coordination around the plan. A detailed overview of this rather complex set methodologies is beyond the scope of the present argument -- in fact, as one project organizer phrased it during an interview, VOI is something that 'most *health economists* don't even understand.' For our purposes, though, the most relevant feature of these models is that they collapse a myriad of heterogeneous parameters -- such as the number of patients to be impacted by an intervention, the economic costs of the diagnostic technologies, the costs of downstream care, quality of life and psychosocial measures, and survival rates -- into a single dollar amount. This pecuniary figure stands in as a summary value of the future knowledge to be gained from conducting further research on a given clinical intervention or scenario around which there is currently some level of uncertainty, premised on an understanding that future decision making will be more informed than at present thus leading to an improvement in clinical outcome and possibly a reduction in expenditures.

One instance where VOI was used came when the team of university-based project organizers -- the group primarily responsible for carrying out the modeling component -- sought to calculate the value of conducting a clinical trial on a set of breast cancer biomarkers. These were one of the initial six diagnostic technology options for a trial which, after several rounds of voting, ultimately surfaced as the highest priority for carrying forward to the trial design phase. As it turned out, there was little prior research addressing the issue of whether the use of these tests improved survival in breast cancer patients; in the absence of sufficient evidence-based findings, project organizers consulted with experts, using their opinions to inform the survival parameter in the model. As a result, they accorded a wide confidence interval within the overall model which was reflected in the diagrammatic presentation of modeling results (cf. Thariani et al. 2013).[12] In this particular case, the values of conducting the trial existed as a pecuniary range, broken down along two parameters that

project organizers understood to be driving this overall value: the number of patients that would be accrued to the prospective study on the one hand, and the confidence in expert opinion about improvements in patient survival which trial results would further inform (Figure 4.1).

Figure 4.1 Expected Value of Sample Information for a trial on breast cancer markers
From (Thariani et al. 2013)

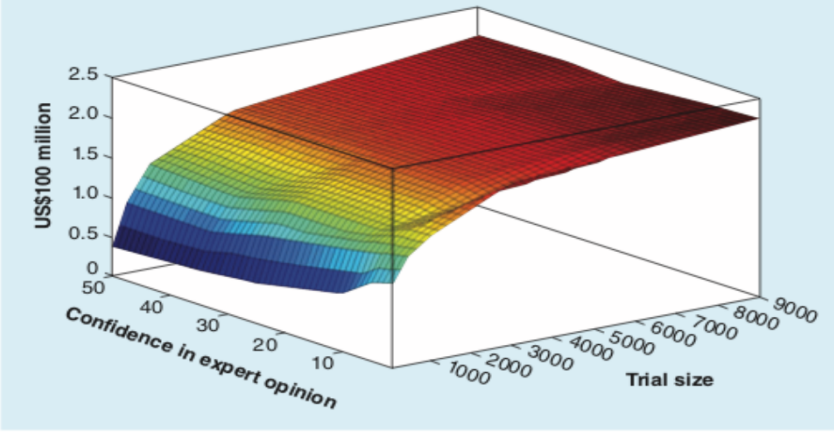


Figure 3. Expected value of sample information values (interpolated fit) as a function of overall trial size and weighting of expert opinion. That is, how accurate we believe experts are, relative to our belief in a trial ranging from n = 1 to n = 50, on survival parameters. The z-axis represents expected value of sample information (in US$100 millions).

In this way, value of information enacted 'the explicitness of alternatives and variables, the rigor of reasoning, and the construction of different scenarios' which, in turn, appeals back to the stakeholders' own preferences and evaluations (Muniesa & Callon 2007:176). More to the point, project organizers deployed these models and their economized outputs as part of a procedure that structured information in such a way that it 'ensured a reasoned response' from the stakeholders (Espeland 2000:1098). When presented with the breast cancer modeling results, each individual stakeholder was placed in a position of having to decide how much confidence they had in expert opinions about survival, which in turn corresponded to a particular value of the trial. Moreover, presented with data which indicated that a trial of 9000 participants was an order of magnitude more valuable than a trial of 1000 patients, a stakeholder who at first preferred the latter had to then decide, based on this new information, whether they still favored a smaller trial or change their mind and opt for the larger -- trial design, which would nevertheless be more expensive to run and take

longer to complete than its more diminutive counterpart.

As functional objects, the VOI models were thus a way of rendering 'knowable options to be chosen' along the path to realizing a single trial to be carried out (Thévenot 2014b:17). More than that, however, VOI models contributed to a particular way of formatting the means for and avenues by which stakeholders and project organizers interacted with each other as they went about executing the prioritization plan. In this way, they functioned as a type of *intermediary object* in the overall coordinative matrix (ibid., cf. Latour 1987; Vinck 2012). The term 'object' here points not only to the 'material support' with which such entities furnish a situation, but also 'designates the focus of human beings' attention or emotion towards which communicating and differing expressions are directed' in an engagement (Thévenot 2014b:17).

There was in fact very little evidence of stakeholders' outrightly questioning the overall process -- in the sense of a mode of *public justification*, e.g. an appeal to higher principles as to how the project was organized. I did, however, encounter a notable exception to this during my fieldwork at the March 2011 ESAG meeting. At a dinner for participants on the evening prior to the meeting, one of the stakeholders, who had been invited to represent the *practicing clinician* perspective, was seated at a table with several of the university-based project organizers who had conducted the VOI modeling. As they conversed, the stakeholder in question had remarked that the modeling was nothing but a bunch of 'statistical malarkey,' signaling his unease with the way that the models reduced prioritization to a set of quantitative, economized, and perhaps even arbitrary numbers (cf. Lampland 2010). This then became a running joke at the meeting the following day, as project organizers took up the term to ironically describe VOI analysis during several presentations in which modeling results were delivered to the stakeholders.

A second moment of differing vis-a-vis intermediary objects arose at an earlier moment in the project when, during the July 2010 meeting, the stakeholders were asked to vote three of the initial six technologies 'off the island,' leaving only the three highest ranked diagnostic tests which would carry forward to the next phase of prioritization. In typical form, the project organizer leading the discussion reminded the ESAG members of the importance of having 'your [own] registered clicker that is in front of you and that you don't swap clickers with anybody else, because they are going to be looking at the results by respondent.' The vote appeared to go off without a hitch: 'Alright,' she

said, as she began to present the results of the exercise, 'now we are going to go through each one -- but which test did you rank first?' As the conversation continued, the organizer paused; there was a glitch in the system, creating some confusion: 'We are sorry about this. This is part of the limitation with the audience response.' One of the stakeholders replies, jokingly: 'Did you say "limitations of the *audience*"'? The transcript records laughter from the group, and a subsequent follow-up from another voice: 'For a democracy, this thing is pretty slow. The Chinese have made a decision and moved on.'

This last comment, once again, highlights the ways in which intermediary objects -- such as the audience voting and display system -- channeled the expression of preferences and options amongst stakeholders during the ESAG meetings, while also acting as a mechanism for channeling their critiques, however subtle and ironic they may have been. It also ironizes on the sometimes uneasy coupling of deliberative methods together with the deployment of technologies that consistently sought to capture the outcomes of those deliberations, quantitatively rendered (e.g. five stakeholders may have ranked technology A as their top priority, while 3 others ranked technology B in the same position.). In fact, as the group went about carrying out the plans of prioritization and design, it was not always clear to stakeholders whether there was parity of importance for project organizers between two distinct objectives in play: the incorporation of their individual knowledge and experiences into the prioritization and design of a socio-technically robust clinical study on the one hand, and project organizers' own efforts to demonstrate the successful implementation of technologies of eliciting stakeholder perspectives and devices for quantifying and representing those preferences (Lezaun & Soneryd 2007; Muniesa 2014). This was in fact a tension that ran throughout the project, rendered somewhat surprising when read in the context of the introductory presentation that one of the project organizers (a bioethicist) delivered at the start of that first ESAG meeting in July 2010. 'This is a power issue,' she professed to the group: 'We want to make sure that we avoid the dominance of technical knowledge over the knowledge and experience of everyone who is involved and engaged and interested in this health research process.'

5. The powers of confinement: a *pre-public* platform

As we have seen in this paper, stakeholder engagement as enacted within CANCERGEN has features of what some have previously described in terms of an 'interest group' or 'interest

representation' model (Espeland 2000, cf. Yearley 2005, ch. 3) but which we have further specified according to a regime of *engaging in plans* and a coincident *liberal grammar* of choices and options (e.g. Thévenot 2007). Organizing participation exercises in such a way works at once to open up decision-making to previously unaccounted for perspectives while in the very same move also *reduces* the absolute number of voices that contribute to making those decisions. Through various qualification processes, the powers of choice are delegated to a handful of human spokespersons tasked with standing in for an entire collective of relevant entities, as was the case with the specific stakeholder categories in CANCERGEN (cf. Akrich et al. 2002; Callon 1984; Ribes et al. 2013). Accordingly, beyond the good, agency, environment, information, objects, and grammars by which we have heretofore characterized planned engagements, we may thus add a still further -- and perhaps more meso level -- element upon which these dispositifs operate. This is what Gilbert and Henry (2012) describe in terms of a *confinement logic*, a particular type of *problem definition* mechanism that has not received at much analytic attention as more *public* settings of problem definition, whose 'mechanisms are assumed to come into play in connection with a space of public debate or through publicity processes that use the media, the legal sphere, or any other domain or mode of public action in which the aim is to broaden the real or potentially concerned or interested public' (43).

Within the more limited spaces of confinement logic, one finds a certain relationality between actors that aims to preserve and balance those relations, striving for routine operation in what the authors call 'an "ordinary" state of affairs' (ibid.:44), analogous to Thévenot's characterization of planned engagements as a regime of "'normal action" or the "normal format" of action' (Thévenot 2007:417). According to this logic, problems are 'defined as a function of multiple priorities' and 'subject to differing choices and compromises' without the requirement to enlarge the generality of one's claims as is the demanded by the regime of public justification (Gilbert & Henry 2012:44). It is when promises and arrangements made in these discreet settings are broken that they tend to overflow, leading to new 'phases' of publicity of the issue. So while in Section 3 we have situated the regime of planning within a wider architecture of regimes that also includes both *public* and *familiar* modes of engaging, an appreciation for the way in which confinement logics operate on planned engagements helps to understand the latter relative to the former two regimes, calling into view a

unique spatio-temporal configuration of planning that is nevertheless related to these other possible modes of engaging with the world.

Let us return, then, to one final and brief interaction that occurred amidst stakeholder deliberations at one of the in-person ESAG meetings. Here, the topic of discussion was one of the shortlisted diagnostic technologies, in this case an assay used for identifying a subset of lung cancer patients who were believed to derive greater benefit from a particular chemotherapeutic regimen based on the genetic profile of their disease. There was broad consensus that a large-scale clinical trial of this test was infeasible and so project organizers and stakeholders had instead floated the idea of conducting a retrospective analysis. Despite some of their methodological limitations, retrospective studies are generally viewed by biomedical researchers a more streamlined way of gathering medical data, and were proposed here as an alternative to designing what would be a lengthy and expensive prospective trial that was uncertain to complete. One way such studies are frequently done is by mining health insurance claims data for intervention being studied, tracking their use by way of health insurance coverage codes (called CPT codes) that are often mapped onto particular classes of tests, thus lending researchers a better sense about how widely a test is being used in practice, for example (e.g. Lau et al. 2011).

The understanding in this particular case was that there was inadequate data about how and why doctors and patients use the lung cancer test in clinical practice, but that such information could be helpful in designing a better randomized trial at some point in the future. Several stakeholders, including those representing the payer perspective, raised the issue that billing codes have not kept pace with more recent technological advances and so tests are instead covered under a range of different existing billing codes, making them poor proxies for tracking usage patterns. Later on in the afternoon, discussion turned to the issue of how CANCERGEN could go about designing a trial for this technology, and what questions could actually be addressed in such a study:

> *[Organizer]*: Some of the points you made, though, about problems with the codes and things like that are solutions that we actually need so that we can actually use those things better than we can currently.
> *[Regulator]*: Exactly. And *to even publicize the challenges with an advantage of this being kind of a stakeholder-engaged process is perhaps -- we can make some noise too around what is holding us up.* Because I think a lot of times it's the research community [who] talks a lot about these challenges and it doesn't make it into a broader

> context. It doesn't reach maybe the ears that it needs to reach. *I'm not volunteering to undergo any public policy agenda or anything like that, but I do think this fits into a broader conversation on how to make system changes.*
> *[Industry]*: CANCERGEN also needs to understand the world we are working in. This is not ten years ago. There is rapid dissemination of information by people who do this as a business. So you have to find a mechanism that allows you to answer these kinds of questions and it seems to me that the natural accompaniment to CANCERGEN and this protocol [...] *could be advertised as a step towards that*, which is to figure out a mechanism that efficiently deals with new technology as it arrives and could conceivably affect the use of therapeutics and treatment in patients.

Surfacing from the regulator stakeholder's comment in this exchange is a reflexive acknowledgment of a horizon of publicity, but only rapidly followed by recourse to the confinement logic of the immediately circumscribed planned engagement. Further describing situations in which such logics are deployed, Gilbert and Henry go on to argue that actors

> decide, exchange information and opinions, negotiate and occasionally agree on this or that decision or public policy direction -- without necessarily taking an official, public position at the same time, and above all without seeking to render those decision-making mechanisms compatible with any public position *they might eventually take*. The fact that these actors are *not required to publicly explain or justify their choices* and the compromises they accept *leads to agreements that could not be reached using similar modes in public spaces*(Gilbert & Henry 2012:43, italics mine).

It is in this sense that I propose that the particular decision-making environment that CANCERGEN typifies (and of which Espeland's own case stands in as an exemplar as well) -- frequently referred to as 'stakeholder engagement,' but to which even some instances of what is otherwise called 'public involvement' may correspond -- constitutes what I call a *pre-public platform*. What do I mean by *pre-public platforms* here? There is, of course, little doubt that CANCERGEN was funded by public grant money; that its concerns were with the health of the American public; and that the trial was to be carried out within the NCI Cooperative Group Program, the publicly administered national cancer clinical trials infrastructure. Yet the decisions reached between CANCERGEN project organizers and stakeholders occurred at a moment *prior* to their

publicization; indeed, certain things that were decided upon within that confined space resisted any form of public issue formation altogether (cf. Marres 2012).

Following Gilbert and Henry, I argue here for the need to shift analyses of participation practices away from discrete 'phases,' instead 'resituat[ing] them in a wider process wherein they are understood as one specific moment *among others* in problem construction' and resolution (2012:49). By extension, this is also an argument for moving beyond dichotomous framings of participation practices as we find when Espeland pits 'democratic' against 'bureaucratic' decision making or Marres' own way, via Dewey, of differentiating 'public issues' from 'private habits' (Marres 2012:48; cf. Lee 2007). In this way, the *regimes of engagement* architecture becomes an especially useful heuristic tool which, re-approached in light of the foregoing argument, does not simply propose the existence of a series of discrete engagement formats but rather occurs as an ecology of engagements and (e)valuations through which actors pass as they go about living in the world, alone and together (e.g. Berger 2015; Brighenti 2016; Star 1995).

Opened up by this approach is the possibility for novel ways of exploring what in the sociological vernacular is referred to as 'power' and 'oppression' as differentiating between regimes 'allows a systematic account of the pressure which a regime puts on another, thus discrediting the confidence underpinning the engagement and damaging the good pursued' (Thévenot 2011:54). The way in which stakeholder engagement was enacted in CANCERGEN is thus indicative of 'the contemporary insistence on autonomy, project, choice and enlightened consent,' all of which 'presume capacities serving engagement in a plan,' and in turn 'put a high degree of pressure' on other modes engagement. That its, engaging in plans strains both the 'localized and familiar attachments' of familiar engagements and, more importantly for the present argument,' also 'eclips[es] the common public' (ibid.). This echoes Dewey's own assertions about the 'inherently problematic' nature of public configurations, which Marres explains as the condemnation of 'communities of the affected' to 'inchoate, obscure, and unstable existences as the kinds of troubling effects that call them into being are likely to remain under-documented in public discourse' (2012:48). Severe constraints are thus placed upon the ability for these communities to transform themselves into 'politically effective' publics (ibid.), as they 'cannot use inherited political agencies' whose institutionalized entrenchment 'obstructs the organization' of new publics (ibid.).

In Dewey's (1946) own words, these institutionalized forms 'prevent that development of new forms [...]. To form itself, the public has to break existing political forms. This is hard to do because these forms are themselves the *regular means* of instituting change' (31). But as we have seen throughout the foregoing analysis, even newly developed forms of participation can be designed in such a way that specifically prevents an escalation to publicity at some future moment. This may well allow certain decisions to be made, and in a way that is more accountable to a greater number of societal actions, but this likely comes at the expense of enabling other modes of expression and of achieving commonality.

6. Conclusion: going 'beyond the public'

I began this paper by highlighting some of the challenges that come along with making sense of particular formattings of engagement, and asking about the tools we might have access to for mapping the terrain of engagement that may lead us out from the 'agora' or the 'hybrid forum' -- indeed, perhaps out from the 'public' itself -- and instead towards some *other* kind of situation. Drawing from an empirical analysis of stakeholder engagement practices in CANCERGEN, in tandem with Thévenot's regimes of engagement framework, I have sought to demonstrate what such an analysis might look like. Beginning with a case in which it is the actors themselves who claim to not be doing 'public' or 'citizen' engagement, but rather 'stakeholder engagement,' is a convenient place to begin pushing the boundaries of the analytic language currently available to students of participation practices. Studies of participation practices in STS scholarship and related literatures tend to deploy and reproduce a vocabulary of 'publics,' 'citizens,' 'patients,' and 'stakeholders' but frequently stop short of analyzing how the designation of these categories of participant relate to overarching coordinative dispositifs and their corresponding qualifications and modes of evaluation (e.g.. Bogner 2012; Lehoux et al. 2012; Nowotny et al. 2001; Ureta 2015). In my view, taking this further step promises to elevate the quality of such analyses such that they may not simply evaluate the successes and/or failures of efforts to involve these groups in decision making, but also that the analyst can articulate at a higher resolution the situational politics which underlie the actors' own criteria for success and failure and how these relate to the particular evaluative good(s) orienting different participation exercises.

To this end, in this paper we have followed the actors' own terms as a means of sensitizing the analysis to a particular enactment of 'stakeholder engagement.' This entailed an examination of the work that project organizers carried out as they went about qualifying what a 'stakeholder' was, as well as qualifying the objectives that their 'stakeholder engagement' was expected to accomplish. Following this, we looked to the *evaluative goods* orienting the engagement, the *agency* of actors, the *environment* in which the engagement occurred, to the *objects* and *information formats* deployed in these situations, and to the particular *grammar* that actors used communicate and differ in their work of prioritizing and designing a clinical study. Through this, it became evident that actors contrasted their own participatory practices to 'citizen' and 'public' engagement in healthcare, which led into some other kind of situation that I have argued constitutes a mode of planned engagement. In this regime of planning, agency takes the shape of a stakeholder; action is oriented towards achieving a stated objective; environments are functionally equipped, lending confidence to the coordination; while commonality is composed through a negotiation of interests, without appeal to a higher order of justice. Finally, we considered all of these elements together with a logic of confinement which collectively characterize *dispositifs* of planned engagement.

These are configured in line with something slightly more demanding than what we find in the case of familiar attachments, yet less demanding than the search for a 'common good' that is required of engaging in public justification (Boltanski & Thévenot 2006). It is a limited coordinative format that can at times occur as oppressive as, for instance, when one attempts to question or critique the exercise on the grounds of a higher order principle (e.g. Cheyns 2011; Silva-Castañeda 2015; Thévenot 2014b). This is of course not to say that there is no place for critique or resistance within planned engagements, but rather that they tend to take much subtler forms and are framed by the available options within deliberations and the mechanisms meant to capture the diversity of perspectives (cf. Berger 2015; Pattaroni 2015). More to the point, however, is that the format itself -- and the decisions reached within CANCERGEN -- are not properly public, but rather exist *prior* to becoming so. Noortje Marres has observed that ' [i]t is far from self-evident that attempts to open up issues for critical outside scrutiny can be qualified as instances of "public involvement,"' going on to argue that

> if we wish to account for the difference that publics can make to politics, we must focus on attempts at the public-ization of issues. It requires that we

> attend to a broad range of events in which issues are articulated as objects of potentially widespread concern. Such an approach, moreover, must acknowledge the real possibility that such attempts fail, and that accordingly no public involvement in politics, in the sense of widespread mobilization of actors, is achieved' (Marres 2007:775-776).

Indeed, it may well be the case that no attempt at publicization is made in the first place.

Returning to the final transcript excerpted in the previous section, what we find is a stakeholder representing the regulator perspective expressing that one possibility is to take the agreements reached within the project and to 'publicize the challenges' and 'make some noise around what is holding us up,' but that he's also not 'volunteering to undergo any public policy agenda or anything like that.' In other words, there is a *public* on the horizon, but it is remains firmly and exactly in its place: on the horizon. It is here, and for this reason, that I argue that CANCERGEN instantiates a form of engagement that is *pre-public*. Situated within a wider *ecology of engagements and (e)valuations*, decision-making platforms like CANCERGEN afford a twofold opportunity for appreciating the variegated and situated *formats* that participation can take, as well as the ways in which these different formats themselves punctuate the flow of time as a series of moments that may qualify for publicity in one moment, but may also call for other modes of coordination which nevertheless imply their own forms of reduction and, at times, oppression. It thus behoves the analyst to take actors seriously when they proclaim, as in the case of CANCERGEN, that their participation practices 'go beyond the public' and to dutifully account for to the type of politics that the confined logics of planning simultaneously afford and preclude.

7. References

Akrich, M., Callon, M., Latour, B., & Monaghan, A. (2002). The key to success in innovation, Part II: The art of choosing good spokespersons. *International Journal of Innovation Management*, *6*(2), 207–225.

Barthe, Y., Blic, D. de, Heurtin, J.-P., Lagneau, É., Lemieux, C., Linhardt, D., … Trom, D. (2013). Pragmatic Sociology: A User's Guide. *Politix*, *103*(3), 175–204.

Beck, U. (1992). *Risk Society: Towards a New Modernity*. London: SAGE.

Bénatouïl, T. (1999). A Tale of Two Sociologies: The Critical and the Pragmatic Stance in Contemporary French Sociology. *European Journal of Social Theory*, *2*(3), 379–396.

Berger, M. (2015). The politics of copresence: an ecological approach to resistance in top-down participation. *European Journal of Cultural and Political Sociology*, *2*(1), 1–22.

Bijker, W. E., Hughes, T. P., & Pinch, T. J. (Eds.). (1987). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT Press.

Blokker, P. (2011). Pragmatic sociology: Theoretical evolvement and empirical application. *European Journal of Social Theory*, *14*(3), 251–261.

Blokker, P., & Brighenti, A. (2011). An interview with Laurent Thévenot: on engagement, critique, commonality, and power. *European Journal of Social Theory*, *14*(3), 383–400.

Bogner, A. (2012). The paradox of participation experiments. *Science, Technology & Human Values*, *37*(5), 506–527.

Boltanski, L., & Thévenot, L. (2006). *On Justification: Economies of Worth*. Princeton: Princeton University Press.

Bora, A., & Hausendorf, H. (2006). Participatory science governance revisited: Normative expectations versus empirical evidence. *Science and Public Policy*, *33*(7), 478–488.

Brighenti, A. M. (2016). Expressive measures: an ecology of the public domain. In K. Avramidis & M. Tsilimpounidi (Eds.), *Graffiti and Street Art* (pp. 119–134). London: Routledge.

Callon, M., Lascoumes, P., & Barthe, Y. (2009). *Acting in an uncertain world: An essay on technical democracy*. Cambridge, MA: MIT Press.

Callon, Michel, & Rabeharisoa, V. (2003). Research "in the wild" and the shaping of new social identities. *Technology in Society*, *25*(2), 193–204.

Centemeri, L. (2015). Reframing problems of incommensurability in environmental conflicts through pragmatic sociology: From value pluralism to the plurality of modes of engagement with the environment. *Environmental Values*, *24*(3), 299–320.

------. (2017). From Public Participation to Place-Based Resistance. Environmental Critique and Modes of Valuation in the Struggles against the Expansion of the Malpensa Airport. *Historical Social Research / Historische Sozialforschung*, *42*(3), 97–122.

Cheyns, E. (2011). Multi-stakeholder initiatives for sustainable agriculture: limits of the 'inclusiveness' paradigm. In S. Ponte, P. Gibbon, & J. Vestergaard (Eds.), *Governing through standards: Origins, drivers and limitations, eds. S. Ponte, S., P. Gibbon and J. Vestergaard* (pp. 210–235). Basingstoke: Palgrave Macmillan.

Degeling, C., Carter, S. M., & Rychetnik, L. (2015). Which public and why deliberate?–A scoping review of public deliberation in public health and health policy research. *Social Science & Medicine*, *131*, 114–121.

Deverka, P. A., Lavallee, D. C., Desai, P. J., Armstrong, J., Gorman, M., Hole-Curry, L., … Veenstra, D. L. (2012a). Facilitating comparative effectiveness research in cancer genomics: evaluating stakeholder perceptions of the engagement process. *Journal of Comparative Effectiveness Research*, *1*(4), 359–370.

Deverka, P. A., Lavallee, D. C., Desai, P. J., Esmail, L. C., Ramsey, S. D., Veenstra, D. L., & Tunis, S. R. (2012b). Stakeholder participation in comparative effectiveness research: defining a framework for effective engagement. *Journal of Comparative Effectiveness Research*, *1*(2), 181–194.

Dewey, J. (1946). *The Public And Its Problems*. Chicago: Gateway Books.

Dodier, N. (1993). Review article: Action as a combination of 'common worlds'. *The Sociological Review*, *41*(3), 556–571.

Ehrenstein, V., & Neyland, D. (2016). *Trials and tribulations of a global health investment: the advance market commitment for pneumococcal vaccines*. Unpublished manuscript.

Epstein, S. (2016). The politics of health mobilization in the United States: The promise and pitfalls of "disease constituencies." *Social Science & Medicine*, *165*, 246–254.

Espeland, W. N. (2000). Bureaucratizing Democracy, Democratizing Bureaucracy. *Law & Social Inquiry*, *25*(4), 1077–1109.

Eulriet, I. (2014). The Civil Sphere and On Justification: Two Models of Public Culture. In S. Susen & B. S. Turner (Eds.), *The Spirit of Luc Boltanski: Essays on the "Pragmatic Sociology of Critique"* (pp. 413–423). London: Anthem Press.

Eymard-Duvernay, F., Favereau, O., Orléan, A., Salais, R., & Thévenot, L. (2005). Pluralist integration in the economic and social sciences: the economy of conventions. *Post-Autistic Economics Review*, *34*(30), 22–40.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. London: SAGE.

Gilbert, C., & Henry, E. (2012). Defining Social Problems: Tensions between Discrete Compromise and Publicity. *Revue Française de Sociologie (English Edition)*, *53*(1), 31–54.

Goffman, E. (1956). *The presentation of self in everyday life*. Edinburgh: University of Edinburgh.

Guggenheim, M., & Potthast, J. (2012). Symmetrical twins: On the relationship between Actor-Network theory and the sociology of critical capacities. *European Journal of Social Theory*, *15*(2), 157–178.

Harrison, S., & Mort, M. (1998). Which champions, which people? Public and user involvement in health care as a technology of legitimation. *Social Policy & Administration*, *32*(1), 60–70.

Holden, M., Scerri, A., & Owens, C. (2013). More Publics, More Problems: The Productive Interface between the Pragmatic Sociology of Critique and Deweyan Pragmatism. *Contemporary Pragmatism*, *10*(2), 1–24.

Horst, M., & Irwin, A. (2010). Nations at Ease with Radical Knowledge: On Consensus, Consensusing and False Consensusness. *Social Studies of Science*, *40*(1), 105–126.

Irwin, A. (2001). Constructing the scientific citizen: science and democracy in the biosciences. *Public Understanding of Science*, *10*(1), 1–18.

Jasanoff, S. (2012). *Science and public reason*. Abingdon: Routledge.

Joly, P.-B., & Kaufmann, A. (2008). Lost in translation? The need for 'upstream engagement' with nanotechnology on trial. *Science as Culture*, *17*(3), 225–247.

Knaapen, L., & Lehoux, P. (2016). Three Conceptual Models of Patient and Public Involvement in Standard-setting: From Abstract Principles to Complex Practice. *Science as Culture*, *25*(2), 239–263.

Lampland, M. (2010). False numbers as formalizing practices. *Social Studies of Science*, *40*(3), 377–404.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Lau, E. C., Mowat, F. S., Kelsh, M. A., Legg, J. C., Engel-Nitz, N. M., Watson, H. N., … Whyte, J. L. (2011). Use of electronic medical records (EMR) for oncology outcomes research: assessing the comparability of EMR information to patient registry and health claims data. *Clinical Epidemiology*, *3*, 259–272.

Laurent, B. (2017). *Democratic Experiments: Problematizing Nanotechnology and Democracy in Europe and the United States*. MIT Press.

Lee, C. W. (2007). Is There a Place for Private Conversation in Public Dialogue? Comparing Stakeholder Assessments of Informal Communication in Collaborative Regional Planning. *American Journal of Sociology*, *113*(1), 41–96.

Lehoux, P., Daudelin, G., & Abelson, J. (2012). The unbearable lightness of citizens within public deliberation processes. *Social Science & Medicine*, *74*(12), 1843–1850.

Lezaun, J. (2007). A market of opinions: the political epistemology of focus groups. *The Sociological Review*, *55*(s2), 130–151.

------. (2011). Offshore democracy: launch and landfall of a socio-technical experiment. *Economy and Society*, *40*(4), 553–581.

Lezaun, J., Marres, N., & Tironi, M. (2017). Experiments in Participation. In U. Felt, R. Fouché, C. A. Miller, & L. Smith-Doerr (Eds.), *The Handbook of Science and Technology Studies* (pp. 195–220). Cambridge, MA: MIT Press.

Lezaun, J., & Soneryd, L. (2007). Consulting citizens: technologies of elicitation and the mobility of publics. *Public Understanding of Science*, *16*(3), 279–297.

Marres, N. (2007). The issues deserve more credit pragmatist contributions to the study of public involvement in controversy. *Social Studies of Science*, *37*(5), 759–780.

------. (2012). *Material Participation: Technology, the Environment and Everyday Publics*. Basingstoke: Palgrave Macmillan.

Moreira, T. (2012). *The Transformation of Contemporary Health Care: The Market, the Laboratory, and the Forum*. New York: Routledge.

------. (2015). Understanding the role of patient organizations in health technology assessment. *Health Expectations*, *18*(6), 3349–3357.

Muniesa, F. (2014). *The Provoked Economy: Economic Reality and the Performative Turn*. London: Routledge.

Muniesa, F., & Callon, M. (2007). Economic experiments and the construction of markets. In F. M. & L. S. Donald MacKenzie (Ed.), *Do Economists Make Markets? On the Performativity of Economics* (pp. 163–189). Princeton: Princeton University Press.

Nowotny, H. (2003). Democratising expertise and socially robust knowledge. *Science and Public Policy*, *30*(3), 151–156.

Nowotny, H., Scott, P. B., & Gibbons, M. T. (2001). *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge, UK: Polity Press.

Pattaroni, L. (2015). Difference and the Common of the City:The Metamorphosis of the Political' from the Urban Struggles of the 1970's to the Contemporary Urban Order. In J. M. Resende & A. C. Martins (Eds.), *The making of the common in social relations* (pp. 141–172). Newcastle Upon Tyne: Cambridge Scholars Publishing.

Rabeharisoa, V., Moreira, T., & Akrich, M. (2014). Evidence-based activism: Patients', users' and activists' groups in knowledge society. *BioSocieties*, *9*(2), 111–128.

Ribes, D., Jackson, S., Geiger, S., Burton, M., & Finholt, T. (2013). Artifacts that organize: Delegation in the distributed organization. *Information and Organization*, *23*(1), 1–14.

Silva-Castañeda, L. (2015). What Kind of Space? Multi-stakeholder Initiatives and the Protection of Land Rights. *International Journal of Sociology of Agriculture and Food*, *22*(2), 67–83.

Soneryd, L. (2015). Technologies of participation and the making of technologized futures. In J. Chilvers & M. Kearnes (Eds.), *Remaking Participation* (pp. 144–161). London: Routledge.

Star, S. L. (Ed.). (1995). *Ecologies of Knowledge: Work and Politics in Science and Technology*. Albany: SUNY Press.

Stavo-Debauge, J. (2012, 15-18 July). La sociologie dite «pragmatique» et la philosophie pragmatiste, une rencontre tardive. Presented at *Pourquoi le pragmatisme ? L'intérêt du pragmatisme pour les sciences humaines et sociales*, Villa Vigoni, Italy.

Stewart, E. (2016). *Publics and their health systems: rethinking participation*. Basingstoke: Palgrave Macmillan.

Stirling, A. (2008). "Opening up" and "closing down" power, participation, and pluralism in the social appraisal of technology. *Science, Technology & Human Values*, *33*(2), 262–294.

Suchman, L. A. (1987). *Plans and Situated Actions: The Problem of Human-Machine Communication*. New York: Cambridge University Press.

Thariani, R., Henry, N. L., Ramsey, S. D., Blough, D. K., Barlow, B., Gralow, J. R., & Veenstra, D. L. (2013). Is a comparative clinical trial for breast cancer tumor markers to monitor disease recurrence warranted? A value of information analysis. *Journal of Comparative Effectiveness Research*, *2*(3), 325–334.

Thariani, R., Wong, W., Carlson, J. J., Garrison, L., Ramsey, S., Deverka, P. A., … Baker, L. H. (2012). Prioritization in comparative effectiveness research: the CANCERGEN experience in cancer genomics. *Medical Care*, *50*(5), 388–393.

Thévenot, L. (2007). The plurality of cognitive formats and engagements moving between the familiar and the public. *European Journal of Social Theory*, *10*(3), 409–423.

------. (2011). Power and oppression from the perspective of the sociology of engagements: A comparison with Bourdieu's and Dewey's critical approaches to practical activities. *Irish Journal of Sociology*, *19*(1), 35–67.

------. (2014a). Engaging with the politics of participative art in practice. In T. Zembylas (Ed.), Artistic Practices: Social Interactions and Cultural Dynamics (pp. 132–150). London: Routledge.

------. (2014b). Voicing concern and difference: from public spaces to common-places. European Journal of Cultural and Political Sociology, 1(1), 7–34.

Ureta, S. (2015). A failed platform: The Citizen Consensus Conference travels to Chile. *Public Understanding of Science*, 25(4), 499–511.

Vinck, D. (2012). Accessing Material Culture by Following Intermediary Objects. In L. Naidoo (Ed.), An Ethnography of Global Landscape*s and Corridors* (pp. 89–108).

Wynne, B. (2007). Public participation in science and technology: performing and obscuring a political–conceptual category mistake. *East Asian Science, Technology and Society*, *1*(1), 99–110.

Yearley, S. (2005). *Making Sense of Science: Understanding the Social Study of Science*. London: SAGE.

Interlude 2

While Chapter 4 focuses on stakeholder engagement within CANCERGEN and elaborates on the notion of *pre-public platforms*, CANCERGEN's work was not limited to participatory exercises alone. Rather, as we saw in Chapter 3, the project was also a locus for innovation in the development of economic modeling and the movement of these methods upstream in the process of prioritizing and designing CER studies.

After earlier theorizing on pre-public modes of engagement and evaluation, I began to question how the aforementioned economic modeling practices related to this wider theme I had noticed playing out in CANCERGEN. Again, in a grounded theory-informed process, I went back and forth between my own ethnographic data and the literatures in the fields of STS and (economic) sociology as a means of grasping precisely what kinds of economic values I was witnessing being produced within the project.

Chapter 5 is thus an attempt to scale up my theory of the pre-public beyond participation, to also think about modes of economic evaluation themselves as taking on this pre-public quality, and laying the groundwork for further analysis of how participatory and quantitative modes of evaluation were ultimately combined in CANCERGEN, and the political consequences of this work.

Chapter 5

Formalizing valuation in clinical research planning: provisional economization and the certification of worth

Abstract

In a report published in 2010, the Institute of Medicine issued a scathing indictment of the myriad inefficiencies plaguing the National Cancer Institute's Cooperative Group Program, the infrastructure charged with carrying out publicly funded cancer clinical trials in the United States. To avert an ever-deepening state of crisis, one of the report's main suggestions was for the NCI to seek out novel methods for streamlining the prioritization and design of clinical studies. This would help ensure that only the most successful candidate trials would be funded, thereby expediting their launch and completion, culminating in the production of useful biomedical evidence about the interventions under study. Given its ability to quantify the value of competing trial options, thereby facilitating comparisons and prioritization across those options, the hybrid economic/decision-analytic modeling approach called value of information analysis (VOI) was proposed as one possible solution to the NCI's woes. This paper describes one early deployment of VOI, which I approach as an instance of formalizing decision-making. Although the ultimate objective of the approach was framed as improving the allocation of scarce public resources, its experimental enactment in fact constituted a mode of *provisional economization* that participated in *certifying the worth* of competing options. As such, in contrast to recent scholarship in economic sociology on the evaluation of public goods, I argue that the pecuniary outputs of VOI were neither 'market' prices nor any other publicly qualified number Rather, they signal a far more mundane economic number whose primary reference was the locally circumscribed planning exercise in which they were constructed.

1. Introduction

This paper is about organizational (in)efficiency, institutional innovation, methodological experimentation, and the nature of economic calculations. At the center of the analysis is value of information analysis (henceforth: VOI), a hybrid economic-decision analytic modeling approach that

aims to quantify uncertainty with an eye towards gathering of additional evidence for making more well-informed decisions in the future. In what follows, I detail the work of an organization called the Center for Comparative Effectiveness Research in Cancer Genomics (henceforth: CANCERGEN) and the sociotechnical practices of project organizers they went about conducting VOI analyses in the context of a wider stakeholder engagement initiative in biomedical research prioritization. In this case, VOI models resulted in the production of a particular set of economic numbers that served to enact the value of multiple competing investment options. These numbers were then fed back as inputs into stakeholder deliberations centered on prioritizing large-scale publicly funded cancer clinical trials of diagnostic technologies.

Approached as a *formalizing practice* (Lampland 2010), the central contention of this paper is that VOI constitutes a mode of what I call *provisional economization*. Through this modeling process relevant realities are reduced to a set of discrete economized outputs that, serving as functional benchmarks, participate in *certifying the worth* of competing avenues for public investment, albeit in a locally circumscribed planning exercise (Thévenot 1995). The general thrust of my analysis is motivated by what I view to be a pervasive if underlying tension within contemporary studies of (e)valuation practices (Helgesson & Muniesa 2013; Lamont 2009) regarding the nature of the correspondence between processes that render phenomena *economic* on the one hand (e.g. Çalışkan & Callon 2009) and qualifications for *market worth* on the other (e.g. Boltanski & Thévenot 2006; Çalışkan & Callon 2010; Stark 2009, 2011).

A key example of this tension is highlighted in Marion Fourcade's work on the grounding of the Exxon Valdez oil tanker in 1989. Over the course of several papers (e.g. 2011a; 2011b), Fourcade details the processes by which appropriate monetary penalties to be imposed on Exxon were adjudicated in the aftermath of the disaster, which decimated the Prince William Sound and a large tranche of surrounding Alaskan wilderness. At the center of her analysis is a technique referred to as *contingent valuation* by which 'economic experts working to estimate ecological damage [...] calculated the subjective value to individuals of the environment spoiled in the spill,' and which was ultimately chosen based on experts' beliefs that it 'had the potential to yield the high level of monetary compensation public officials were looking for -- a compensation that would properly deter future potentially harmful behavior by Exxon and "make the public whole"' (2011a:1727, 1759). Reflecting

on contingent valuation as one instance of a wider class of technologies of economic valuation that are nevertheless taken up 'in cases where markets do not exist and are not even sought' -- as was the case with the absence of a market for pristine Alaskan wilderness -- Fourcade argues that the 'ultimate outcome' of such techniques has been a granting of increased 'legitimacy and authority' to the 'market logic':

> The reason is since at least World War II, economists have been no less central to the valuation of *non-market goods* than they are to framing the processes that generate *market prices*. Economics operates by and large as if non-market goods were priceable (since they have subjective value, and price and value are deeply intertwined), that is, *as if they were being traded on markets* (2011b:46, italics mine).

Whether the link between the *economic* and the *market* is as clearly defined as Fourcade presents it in her analysis has been the subject of a subtle but important debate, however (Centemeri 2015; Stark 2011; Thévenot 2012).

Laurent Thévenot's rereading of the Exxon case, for instance, claims that despite striving to simulate a 'fictive market price,' contingent valuation methods are in practice more akin to survey instruments that aim to capture individual *citizens' opinions* (Thévenot 2012:5). On the one hand, then, this operation of the survey has elements of a specific mode of communicating in public that he labels the *liberal grammar*, wherein individuals communicate by way of 'choosing among options also accessible to other individuals,' and compose commonality through the process of negotiating among those 'publicly recognized' preferences (2011:199). On the other hand, it is also suggestive of a 'civic' test of worth, which demands for 'individual voices [to] be expressed in a vote' -- as opposed to having those same individuals 'choose, as *customers*, to buy a *market good* at a certain *price* in a situated *transaction*' (Thévenot 2012:5; cf. Boltanski & Thévenot 2006). Thévenot concludes that the 'money equivalence obtained by contingent valuation' in Fourcade's case is thus '[f]ar from a pure technical device' but rather consists of a 'compromise' that stabilizes principles from 'civic' and 'market' worlds within the survey instrument, while being expressed through the liberal grammar (ibid.). Along similar lines, David Stark's own analysis of Fourcade's work claims that it 'invite[s] some critical reflection,' as he suggests that there are in fact 'some good reasons' for not equating 'all monetary valuations' with 'market pricings' (Stark 2011:321). Looking to monetary valuation under

state socialist central planning, for instance, Stark argues that it is 'not without reason' one may argue that these valuations 'were not really prices' but were rather 'some *other kind of numerical assignment* but not really money' (ibid., italics mine).

In Stark's view, questions such as these are directly linked to a wider set of questions about how accounting calculation practices are grasped: 'If we posit that there are non-market orders of worth,' he ultimately asks, 'do these have distinctive rationalities that are nonetheless recognizable as "calculation"' (ibid.)? The present paper ventures to provide an affirmative answer to this question through a close empirical analysis of the rollout of value of information analysis in CANCERGEN. In doing so, I draw from Lampland's (2010) delineation of two types of numbers that she has encountered in her own studies of calculation and accounting. A first type, which she labels *provisional numbers*, are those 'used in planning and strategizing' and work to 'assist groups in setting the parameters for tasks at hand and debating their relative merit' (378). With *false numbers*, a second type she specifies, 'the primary task is to learn how to deploy numbers, making the relative accuracy of the numerical sign less important than the attempt to master the logic of formal procedures' (ibid.). Situating monetized VOI model outputs as simultaneously 'false' and 'provisional,' I argue that VOI is exactly such an economic assemblage with its own 'non-market' calculative rationality.

I begin in section two with a discussion of the wider setting in which VOI emerged. Here, I highlight the growing acknowledgement of the myriad inefficiencies imperiling the publicly-funded national cancer clinical research infrastructure in the US during the first decade of the 21st century; and the subsequent establishment of the Center for Comparative Effectiveness Research in Cancer Genomics (henceforth: CANCERGEN), an organization that positioned itself as a site for experimenting with value of information methods as a means of streamlining and formalizing the prioritization and design of large scale cancer clinical trials. In section three, I describe the emergence of VOI in CANCERGEN, focusing on the justification for its use in research prioritization and its affordances as a technology of *commensuration* (Espeland & Stevens 1998). The fourth section goes on to explore the series of *calculative* and *cognitive* strategies that project organizers carried out as they went about introducing the approach to a group of external stakeholders with little prior background in VOI methods. This, then, brings us to the penultimate section, where I zoom in on the particular strategy of presenting VOI in a specifically *economized* information format,

suggesting the dual ontology of these numbers (e.g. Mol 2002) that are at once *provisional* and *false*. Using this ontological doubling, I conclude with a discussion of how we might approach certain economization practices and their outputs not as publicly-qualified *market* numbers, but rather functioning as locally circumscribed exercises which aim at a more functional *certification of* worth, and some reflections on what this difference makes for social studies of (e)valuation and formalization.

2. Formalizing prioritization

In early March of 2014, the US National Cancer Institute rolled out its National Clinical Trials Network (NCTN): a state of the art organizational initiative that emphasized industry partnerships and the use of next-generation DNA and RNA sequencing technologies to achieve precision (or 'personalized') medicine in oncology at the national scale. The NCTN would come to replace NCI's existing multi-disciplinary and multi-institution network of researchers known as the Cooperative Group Program (CGP), which since the 1950s had been the primary infrastructure for conducting large scale publicly funded cancer clinical trials in the US (Keating & Cambrosio 2012).

This significant restructuring of the national clinical trials system followed a period during which stakeholders within and outwith the NIH and clinical research communities brought the (in)efficiencies of the CGP under increasing scrutiny. Many of the concerns that these groups expressed are reflected in a landmark Institute of Medicine report (Nass et al. 2010) that outlined a series of challenges the CGP found itself confronting as the first decade of the 21st century came to a close: 'Stagnant funding, inefficient processes, extensive and complex government oversight, and a growing trend toward the conduct of industry trials overseas have contributed to inadequate physician and patient participation in clinical trials, threatening the Cooperative Group Program's ability to efficiently translate discoveries into clinical applications' (64). Left unaddressed, the already 'terrible waste of human and financial resources' would only worsen, driving the CGP into an ever deeper 'state of crisis.' To ensure its survival, the report issues the Program an injunction to 'streamline oversight, […] select and prioritize trials more stringently, fully fund the most innovative and promising studies, and open and complete trials with greater speed' (ibid.:x).

Although many of these objectives would ultimately come to inform the NCTN's updated organizational structure, the years preceding this transformation witnessed a set of more distributed,

ad hoc, and experimental attempts at achieving organizational efficiency in cancer clinical research. In early 2009, for instance, the National Cancer Institute issued a new funding mechanism that it called Research and Research Infrastructure "Grand Opportunities" (or GO Grants), which included one particular stream of funding to support 'Comparative Effectiveness Research in Genomic and Personalized Medicine.' Comparative Effectiveness Research (CER), a nominally new type of health outcomes research, was a significant budget line in the so-called Stimulus Package that president Barack Obama signed into law that same year, receiving over US$1 billion of federal support (Hoffman et al. 2016). At the same time, there was also a rapid uptake of genomics and personalized medicine (GPM), where novel tools driven primarily by high throughput -omics technology were presenting an accordingly novel set of technical challenges and regulatory uncertainties in both clinical research and clinical practice. The need for improved efficiencies in clinical research was thus seen to be all the more pressing, with the text of the latter CER/GPM Request for Applications urging the convening and deployment of multi-disciplinary teams that 'would function to develop multi-level and systems approaches to build CER capacity and tools in GPM, and define a new future for CER using accelerated timelines to transform clinical and public health practice' (NCI 2009a).[2]

The Center for Comparative Effectiveness Research in Cancer Genomic Medicine (CANCERGEN) was one of seven programs that the NCI ultimately funded under the this scheme. Led by an MD/PhD Principal Investigator with expertise in clinical research and health economics, and based at a large academic cancer hospital in the Pacific Northwest, CANCERGEN's guiding vision was to 'overcome structural barriers that have limited the quality and timeliness of evaluations in GPM, with the result that promising technologies in the cancer genomics pipeline can move rapidly from "proof of principle" to improving the effectiveness and cost-effectiveness of cancer clinical care' (NCI 2009b). This entailed deploying an innovative assemblage of 'expertise in decision

[2] The RFA defines comparative effectiveness research as the 'rigorous evaluation of the impact of different options that are available for treating or preventing a given medical condition for a particular set of subjects' and defines genomic and personalized medicine (GPM) as 'the applications of genome-based approaches in cancer care and prevention through the use of inherited and somatic biomarkers for diagnosis and prediction and drug and other interventions based on these tools.' Further explication of CER is beyond the scope of the present paper, but see Hoffman et al. (2016) for a more thorough conceptual history of this research and its relationship to other fields of healthcare evaluation.

modeling, database linkage, ethics, policy, and clinical trial design [...] to facilitate the rapid design and implementation of prospective CER studies of GPM technologies' (ibid.). Of particular relevance to the discussion at hand, CANCERGEN proposed to use SWOG[3] -- one of the largest of the NCI Cooperative Groups -- as a testbed for conducting the trial(s) generated through its experimental research prioritization and design process.

On the ground, there were two prominent features of CANCERGEN's work. The first was carried out in partnership with a nonprofit organization specializing in stakeholder engagement in the healthcare sector, which worked to facilitate collaborations between CANCERGEN project organizers and a group of 'multidisciplinary stakeholders' and to 'develop methods for identifying CER topics in cancer genomics that are priorities from multiple stakeholder perspectives' (ibid.). Members of this stakeholder panel, dubbed the External Stakeholder Advisory Group (ESAG), were recruited to represent the perspectives of various categories of decision-maker within the healthcare system: patient and consumer advocates; regulators; payers; and practicing physicians, among others. As the lead on this stakeholder engagement component explained to me during an interview, what was interesting for project organizers was to see

> the interface between having a research agenda that's predominantly science-driven versus one that's now being shaped by an external stakeholder advisory group. I think that's sort of the question there. It means that the premise of comparative effectiveness research is, we want to have research that's going to meet the information needs of decision makers in the real world - patients, practicing clinicians, guidelines developers, payers, policy makers. We want to build their information needs into the prioritization and the design of studies and therefore - theoretically at least - if they get involved early, the results [of the study the helped to design] would be more relevant. So [CANCERGEN] is essentially a way to test that.

Moreover, project organizers viewed this objective as falling in line with the aforementioned IOM report, as the CANCERGEN Principal Investigator professed to the project organizers and stakeholders who had convened at the first in-person ESAG meeting in June 2010:

> I would like to say we are visionary because what happened in the interim after we started this is, the Institute of Medicine came out with their report

[3] SWOG was until 2011 known as the Southwest Oncology Group. Despite being founded in Texas in 1956, which garnered it its regional name, the organization had nevertheless been headquartered since 2005 at the University of Michigan in Ann Arbor.

> on the clinical trials system. It was a somewhat critical report [...] and right in the middle of their introduction they talk about the importance of multi-stakeholder input. In the middle there, it says all participants and stakeholders - including physicians, patients, healthcare insurers as well as the NCI and regulatory agencies - must reevaluate their contributions to this cooperative group process. This, I think, was serendipitous for us, but it really anticipated what we are all about in this research project. [...]

A manuscript detailing CANCERGEN's approach to stakeholder engagement, written by several project organizers and published a couple of years into the project's funding cycle, describes adopting a method that 'relied on *qualitative* evaluation of the potential value of research,' where they gathered 'diverse stakeholders' with 'unique perspectives' in a 'neutral forum' to support 'open discussion' and a 'constructive exchange of ideas and concerns' wherein stakeholders' could collectively examine their 'specific interests and perspectives' (Thariani et al. 2012:392).

During this earlier qualitative evaluation process, project organizers presented the stakeholder group with a list of six different diagnostic technologies -- most quite novel, but a few older ones as well -- that were understood as being routinely used to monitor disease and guide therapy recommendations in the clinical management of cancer patients.[4] Yet the tests were also problematized in the following way: due to a series of historical and contemporary regulatory contingencies, the tests had proliferated clinical practice without being submitted to the rigors of large scale randomized controlled trials (RCT). Thus, there was no 'gold standard' evidence (cf. Timmermans & Berg 2003) showing that their use led to improved patient outcomes over and above existing standards of care -- what is generally referred to as a technology's *clinical utility*.[5] As one project organizer explained, the the main question they were interested in answering was in terms of comparing these newer or understudied diagnostic technologies against existing 'standard of care' interventions: 'Is it worth pursuing a study [...] comparing this new test to standard of care?' In this

[4] To identify the six candidate tests, project organizers developed a method they called 'landscape analysis' that combined a formal review of published and grey literatures on the one hand, and 'domain-specific expert consultation' on the other, focused on identifying 'interventions at the clinical horizon' and those 'still in the research phase' that would be potential high-value candidates for conducting large scale clinical studies (Thariani et al. 2012:389).

[5] In other words, despite the proliferation and regularized deployment of these tests, the extent to which they produced improved outcomes among cancer patients remained uncertain according to the tenets of evidence-based medicine and its privileged evidence-generating technique, the large scale randomized controlled trial (Marks 2000).

way, the trial would be more than '*just* a test' of its efficacy, but would consider a host of additional features including its 'clinical utility [and] the incremental clinical utility and costs of a new test compared to standard of care.' The task for the stakeholders in this process was to contribute their particular perspectives, understood as being representative of the various decision-making constituencies they were invited to participate on behalf of, as inputs into deliberations about the relative merits of conducting a large scale RCT on each of those six tests. The expected output of that process would then be a consensus-backed, rank-ordered list of the six tests -- from highest to lowest priority -- delivered back to the project organizers, and from which the three highest-ranked tests would, in turn, serve as inputs into the subsequent evaluation phase (and, ultimately, the design and execution of that study.)

It was at this point that value of information analysis came to take center stage in CANCERGEN, for no sooner that the authors of the aforementioned publication admit to using qualitative methods, they immediately go on to announce that they were also 'currently working on *formal* value of information (or value of research) calculations to derive quantitative estimates for the top 3 candidates' (Thariani et al. 2012:392, italics mine). As it would turn out, then, stakeholder engagement was not really *all* that CANCERGEN was about.

3. Devising commensuration

With roots in engineering research and environmental risk management dating back to the 1950s, the development and application of value of information analysis for health-specific use cases has seen extensive elaboration over the last two decades -- in particular by a group of health economists in the United Kingdom, and more recently spreading to many other jurisdictions as well (Tuffaha et al 2013, 2014). According to Claxton and Sculpher (2006), two of the UK's most prominent VOI methodologists, the approach

> values the additional information generated by further research in a way which is consistent with the objectives and the resource constraints of health care provision [...] This allows a comparison of the potential benefits of further research with the costs of further investigation, a comparison and prioritisation of alternative research recommendations, as well as an assessment of the value of investing resources in research or other activities

> [...] In this sense it provides a unified and coherent framework for prioritisation of research and the use of healthcare technologies (p. 1056-7).

The approach entails the production of multi-parameter models whose inputs are gleaned from previously published clinical trial data and, in cases where the relevant clinical data is unavailable or nonexistent, from expert opinion as well. Outputs of the models take the shape of a standardized quantified metric that serves as a measure of prospective future benefits to be gained from conducting further research on a given intervention or clinical situation. VOI can also be used to isolate especially uncertain elements or parameters of a medical intervention or clinical scenario which then recursively inform the outcome parameters to be measured and accounted for in designing and conducting additional research. The posterior (read: projected) value of the intervention that follows from this improved knowledge about its use (in terms of quality/quantity of life or cost, or both) is then be compared to its use under current conditions of uncertainty.

The concept of 'value' is operationalized here in terms of either improved patient outcomes or improved cost-effectiveness in healthcare. Valuable research is thus conceived as that which achieves either or both of these goals through improvements in how patients are clinically managed at some future time. The value of information is a function of four key elements (Bennette et al. 2016):

- the probability of error, e.g. the probability that a wrong decision will be made based on existing information;
- the opportunity costs of error, e.g. the consequences of making a wrong decision in terms of healthcare costs, a patient's life expectancy, or their quality of life;
- the total impacted patient population likely to face the decision in the future;
- and how much new information a given trial will collect.

It works to account for the probability of a change in decision-making as well as the magnitude of the benefit derived from the research itself. Further research is believed to be most valuable when there are high levels of uncertainty around the harms and benefits of a given approach to clinically managing a particular condition, which makes the uptake of the improved knowledge more likely than if there were an already substantiated body of knowledge informing existing clinical management (Wong et al. 2012:1118).

Proponents of VOI view its affordances as twofold. First, it provides decision-makers with a tool for evaluating the costs and benefits of pursuing a single research investment option. The outputs of

VOI models arrive in the form of a standardized quantitative metric that stands as a summary measure of the overall benefit of that investment option in terms of clinical gains, which may also be monetized (more on this below). This in turn enables decision-makers to compare that projected monetized value of a trial with the projected economic costs of conducting the research, such that if the value of the trial outweighs the cost of running it, it is considered a worthwhile research investment. But in providing such a standardized output, this first affordance also translates into a second: by conducting VOI analyses for multiple, possibly competing, avenues for research investment, the standardized output for each option facilitates comparisons of value across the multiple research options and thus enables prioritization of those options based on which is projected to provide the *greatest* overall returns and, by extension, which is also the least valuable of the lot. In sum, VOI is framed as providing organizations and decision-makers with a set of tools with which they can adjudicate the allotment of limited resources for conducting biomedical research.

I first encountered VOI quite early on in my fieldwork with CANCERGEN. During an interview with a project organizer focused on the stakeholder engagement component, she described the various processes and frameworks she and her team had been working on and then mentioned matter of factly that a team of project organizers based at another institution were in the process of developing decision models as an additional element of the prioritization process. Given my own (admittedly biased) expectation that the project was primarily an exercise in broadening stakeholder participation in research decision-making, this struck me as somewhat of a surprise. I asked her to elaborate on the function of these models in terms of CANCERGEN's broader set of aims, to which she replied: 'We decided to have more *formal quantitative assessments* of what would be the value than, if you will, a *qualitative* assessment [by the stakeholders]' and that 'the output of these things is going to be in dollars.' In other words, she explained, in a way similar to how other kinds of decision-economic modeling practices in healthcare 'allow you to compare treatments in terms of [...] [a] common metric,' value of information analysis 'will allow you to compare the value of research in dollars. So it allows you to rank across and say, "Here's where, [...] for a given investment, you would get a bigger value."'

From this foregoing description of value of information analysis, we are able to glean several important and interrelated features of this approach (Table 5.1). In this concatenation of characteristics we find VOI framed as a *formalization practice*. Formalization is a ubiquitous feature of contemporary society and appears under many guises (e.g. Busch 2011; Lampland 2010; Thévenot 1984) -- in this case, enacted in and enabled through value of information analysis, understood here as a particular valuation machine (Dussauge 2016) that aimed to further rationalize decision-making vis-a-vis the provisioning of scarce public economic resources for conducting large scale cancer clinical trials.

Table 5.1 Stepwise features of value of information analysis

- VOI is qualified as being a method of *formal* calculation;
- Calculations produce *quantitative* estimates of value;
- Quantitative estimates are formatted as *economic* numbers;
- Economic numbers serve as a *common metric*;
- Common metrics allow for *comparing value*;
- Comparisons of value enable *ranking* across a range of phenomena;
- Ranking facilitates the *prioritization* of the highest value investment options

One operative mechanism of this valuation machine is *commensuration*, which Espeland and Stevens (1998) describe as 'the expression or measurement of characteristics normally represented by different units according to a common metric' (314). Commensuration is a 'symbolic, inherently interpretive, deeply political' process that is central to 'how we categorize and make sense of the world' but, in doing so, also 'changes the terms of what can be talked about, how we value, and how we treat what we value' (ibid.:314-15). It is also, in Porter's (1998:45) view, one of the 'vital elements' of what Daston (1995) has labeled *mechanical objectivity*: a particular form of objectivity that 'strives to eliminate all forms of human intervention in the observation of nature, either by using machines [...] or by mechanizing scientific procedures, as in deploying statistical techniques to choose the best of a set of observations' (19; cf. Porter 1995). Thusly endowed with such objectivity, numbers are 'grounded simultaneously in an epistemology of authenticity -- in a yearning for "data" in its root

sense of "givens," bestowed with the effortless of grace -- and also in the guilty conviction that fallen humans, left unsupervised, can only meddle with the givens of nature' (ibid.).

The virtues of commensuration therefore include its ability to standardize as *quantities* phenomena whose *qualities* might be otherwise murky or uncertain, which also implies 'condens[ing] and reduc[ing] the amount of information people have to process, which is useful for representing value and simplifying decision-making' (Espeland & Stevens 1998:316). Decisions are thus often made based upon a myriad of parameters and tradeoffs between them which, through commensuration, can be condensed into a series of single summary measures which facilitates the drawing of comparisons. Whether we are to understand VOI analysis and its economic outputs as cost-benefit ratios or prices, two paradigmatic examples of monetized commensuration -- or else as some *other* kind of economic number, in the sense suggested by Stark (2011b) -- is a question I take up later in this paper. But the important point for now it simply that it is the production of specifically *economic* numbers that drives the commensuration process in value of information analysis.

Here, Çalışkan and Callon's (2009) notion of *economization* becomes relevant, a term they use to 'denote the processes that constitute the behaviours, organizations, institutions and, more generally, the objects in a particular society which are tentatively and often controversially qualified, by scholars and/or lay people, as "economic"' (370). The inclusion of the suffix '-ization' highlights the active work of economics: 'that the economy is an achievement rather than a starting point or a pre-existing reality that can simply be revealed and acted upon,' one that is, importantly, driven by particular economic theories (ibid.). As Çalışkan and Callon remind us: 'The economic valuation of things implies their material characteristics are taken into account by agents endowed with judgment and an ability to perform calculations' (2009:384)

To be a calculating agent, however, one must have the necessary equipment with which to perform economic calculations. Latour and Callon (1997) describe one such possible list of phenomena, which includes the drawing up of a list of possible future states of affairs; the establishment of a hierarchy between those states; and the identification and description of actions which will likely result in any of those states. Individual actors, firms, and states so 'deprived of the means to classify, compare, and evaluate' different products or investment decisions 'cannot make but an ill-informed decision' and risk being 'doomed to both impotence and paralysis' (ibid.: 5).

They go on to claim that 'the obsession of economists is to produce the economization of the world by constantly replacing the non-economic, or costly, overflows with relations that are economic in all of the senses of the world' (ibid.:7). Yet not all economic concepts arrive as immediately tractable to their (perceived) users; instead, significant investments are often required to ensure a mutual fit between a (standard, formalized) model and the local circumstances in which it is deployed, as well as the extensibility of formalisms in time and space.

Thévenot approaches this issue in terms of 'investments in form' (1984), signaling the work required for 'producing standard forms which permit the establishment of equivalents -- i.e. the stable and economic articulation of these forms' (ibid::8). His articulation of 'economic' here is not intended to be of the pecuniary type -- or at least not exclusively so -- but rather concerns the entire economy of formalization, standardization, and the production of standard forms as guarantors of coordination. The crux of his argument is that relying on pre-existing standards and categories entails a 'saving[s]' while, to the contrary, consistent recourse to the 'personal interpretation' of information

> may be expensive and is difficult for the producer to standardize. In this way, the criterion from which can be justified on logical principles also proves to be very economical because it makes possible articulations with other fixed forms and avoids the costs inherent in the personal interpretation of such articulations and in the collective negotiations which are needed to achieve uniformity among individual interpretations (ibid.:4-5).

Read in the context of Daston's 'fallen' and 'unsupervised' humans, it appeared that, at least in the eyes of CANCERGEN project organizers. stakeholders' own qualitative evaluations of research priorities were a necessary but not wholly sufficient means for capturing the value of clinical research studies to be conducted within the setting of the NCI's Cooperative Group Program. This recognition of stakeholders' bounded cognitive capacities served as the primary justification for the introduction of formal calculations into the CANCERGEN prioritization assemblage (Boltanski & Thévenot 2006; Çalışkan & Callon 2009).

4. Informationalizing valuation

4.1 From VOI to the *value of research*

Having initially approached my fieldsite as a means for exploring sociological themes of *participation* in biomedical research decision-making, I was struck by an ethnographic strangeness (e.g. Neyland 2008) upon discovering the growing importance of economic modeling in CANCERGEN. I would soon come to realize that the value of information analysis component -- headed up by a team of health outcomes and public health genomics researchers with expertise in health economics -- was a novel implementation of VOI and functioned as something of an a an in vivo experiment in economization (e.g. Muniesa & Callon 2007). As I probed about this during my early interview with the stakeholder engagement lead, my respondent seemed to pick up on and validate my sense of strangeness: 'That was sort of it,' she replied, 'it's brand new!' and proceeded to sketch out a bit more about the project's timeline and process:

> We had the [first] in-person meeting [...] and [the stakeholders] did voting there, and then they did voting following that meeting. [...] Now, we're going to have the [value of information analysis component]. We're going to actually have a teleconference where they vote again based on the [VOI] results. [We're] saying: 'Okay, [given] everything that you've heard *plus* the [modeling] results, would that change your ranking?

A subsequent interview with the Principal Investigator of CANCERGEN confirmed the novelty of the approach: 'To our knowledge it had never been used in the US to make priorit[ization] decisions -- certainly not within a clinical trials research group like we were focusing [on]. So, that was -- it was really completely new territory for us.' Given the institutional context in which CANCERGEN situated itself, where much of its work was oriented towards developing mechanisms for improving the efficiencies of research prioritization and design within the NCI Cooperative Group Program -- discussed previously in Section 2 and which the PI alludes to in the above quote -- VOI analysis found a champion in CANCERGEN. In step with the recommendations put forth in the aforementioned IOM report, project organizers found in this approach a device that

> might be able to give some more information about the value of these tests [...] [W]e were struggling with how to convey to this clinical audience the potential for economic–slash–clinical value, [which] would be beyond what they typically focus on - which is very narrow, intermediate endpoints.

As the PI explains in this interview excerpt, VOI analysis offered a particular calculative rationale that could push beyond the characteristically limited *clinical* perspective -- or what the stakeholder engagement lead had previously referred to as the *science-driven* approach to choosing studies -- that guided how trials had typically been prioritized and designed within the Cooperative Group Program. Project organizers problematized this latter approach as it depended on individual investigators' more personal academic affinities and perhaps idiosyncratic interests in fairly narrowly defined clinical questions. The 'costs inherent in [...] personal interpretation of such articulations' (Thévenot 1984:4) had thus long worked to eclipse precisely what VOI now promised to achieve for SWOG, and for the Cooperative Group Program writ large: a more formal approach to prioritization that, in producing standardized quantitative metrics, could foster shared understandings of a trial's value and thereby streamline coordination in bringing only those most valuable trials to term.

How best to convey this quantitative information was a matter of concern and a source of uncertainty for project organizers. This was amplified only further by the singular opaqueness of VOI, a point the project PI went on to highlight in his recounting that project organizers

> didn't know how to take the VOI information, which was very technical and complicated, and translate it [...] So we spent a lot of time thinking about how we could just get [the stakeholders] to understand what VOI is. And VOI is complicated. Frankly, a lot of economists don't understand what VOI is, let alone lay people or clinicians.

The team thus set out to render the modeling exercise tractable to this group of stakeholders whom, on the whole, were not experts in the decision-analytic methods undergirding the VOI approach and many of whom had little knowledge of health economics. In the course of doing so, project organizers made a number of normative decisions about how they would go about framing, calculating, and presenting model results.

One such decision was about how to *frame* value of information analysis when first introducing it to the stakeholder group. For instance, during an interview with a project organizer involved in

carrying out the VOI component, I questioned whether there were certain preparations that went into presenting this information to the stakeholders. In one part of his answer, where he was explaining some technical details about how VOI can be used to calculate the value of trials in multiple different ways (more on this below), he began tripping over his words, pausing to admit to me: 'These terms are horrible, so we kind of changed it [...] we changed the names around to make [it] a little more intuitive.' There was, in other words, a concerted effort to avoid the 'technical' terms *value of information* and *VOI* -- as I have used them throughout this paper -- and to instead use *value of research* or *VOR* as the terms of art when referring to the modeling approach within CANCERGEN.

4.2 Tractable calculations

But in addition to this shift in nomenclature, my respondent also emphasized that 'you can't just go in cold. You have to educate [the stakeholders] about the approach -- you have to tell them what you're doing, and we did that along the way.' One aspect of this education component entailed a number of conference calls that project organizers hosted to help bring the stakeholders up to speed on VOI analysis. Another project organizer explained during an interview that 'the challenge is that it takes time to do this and the concepts are not easy.' She went on to recount a conference call that project organizers held with the stakeholders to review the models:

> We tried to do a lot of prepwork and to think about -- we went over the slides multiple times, and I kept saying [to the other project organizers]: 'There's too much. You can only pick so many acronyms that people are going to understand about this, so what's the big point you want to get across first in understanding this concept?'

The core concepts driving VOI were further outlined in a 'Value of Information Analysis Briefing Paper' (Carlson et al. 2013, Appendix B) that project organizers developed and circulated among members of the stakeholder group as another component of their educational strategy. The 'Background and Rationale' section of the document, for example, explains that VOI is

> particularly useful when the approach to patient care is very uncertain and when the consequences of making the wrong choice are large, both in clinical and economic terms. VOI can focus research prioritization discussions by highlighting critical information and areas of great uncertainty in the clinical management pathway. In this way, VOI is a tool that can help decision

> makers maximize the impact of research portfolios on medical care and human health (ibid.).

A corollary to this stakeholder education component was a second normative decision about how the project organizers would actually go about calculating the value of information models at this stage of the project. Here, project organizers pursued an approach called *expected value of perfect information* (EVPI), which they explain to the audience of stakeholders under a section of the aforementioned educational backgrounder document entitled, 'The Theory':

> If decision makers had perfect information about the risks, benefits, and cost impact of a particular technology, they would always be able to make correct choices regarding the use of the technology. The right drug or test would be given to the right patient at the right time. *The difference between the value of having perfect information and the value of current information (with attendant uncertainty) is known as the value of perfect information (VPI).* VPI can be interpreted as the maximum amount we would be willing to spend to learn about the benefits and risks of a particular test or therapy. Although having perfect information is not possible, *VPI is the upper level on the value of further research and as such can serve as an initial threshold to support research funding decisions* (ibid., italics mine).

A member of the team of project organizers conducting the VOI analyses explained to me that

> the actual VOR calculations you set up in a way that you say, 'Well, what are all the major pieces of pertinent information that inform the scenario? What if you had perfect information on all of that?' So in a way, expected value of perfect information is kind of all encompassing. In a way, it's a topic.

He emphasized that by using EVPI, they were effectively doing 'the simpler version of VOR' that assumes you are doing an 'infinitely sized' and thus 'perfect trial':

> It's somewhat useful, but its not precise to a specific trial design. We also didn't have - this is an important point - we didn't have any trial designs, we just had topics. They were ranking topics - areas. We weren't saying [...] 'Here's a trial this size with these outcomes.' It was, 'Here's the topic of breast cancer tumor markers.'

For the purposes of clarifying this foregoing description of EVPI as producing a 'topic,' and the significance of project organizers having used this 'simpler version,' we can compare it with another

possible method of calculating VOI -- referred to as the *expected value of sample information* (EVSI).[6] This second, more granular, approach calculates the differential value of specific model *parameters* rather than simply an *upper bound value* based on an infinite sample size and all possible relevant parameters. Researchers use EVSI data, which are produced using probabilistic modeling based on the existing stock of biomedical research data and expert input, to pinpoint the greatest areas of uncertainty in a given clinical decision-making scenario -- for instance, whether it is questions of quality of life, overall survival, the technical performance of a given diagnostic test, and/or some other phenomenon that is driving uncertainty in clinical decision-making. Those VOI outputs can then recursively be used as inputs for the design of further clinical research studies, which (1) parameterize the uncertain area(s) as trial endpoints to be tested, with the aim of redressing uncertainty around them while also (2) enabling researchers to make sample size calculations such that the trial can accrue an optimal number of patients, thus ensuring statistical power of the study and the reaching of durable conclusions about the parameters under investigation (cf. Thariani et al. 2012).

Returning for the moment to EVPI, however, recall that the initial setting for this calculative work occurred at the point where the stakeholders had selected their top three candidate tests. The team of began by calculating EVPI for each of these tests, and their first step was to develop three decision-analytic models, one for each of the diagnostic tests under consideration. Each model was populated with a number of different variables reflecting 'key events' that occur in the trajectory of the disease and testing strategy in question (e.g. in lung and breast cancer), beginning with the use of the diagnostic test in question and carrying through to the final health outcomes, including death. Variables were parameterized along axes such as life expectancy; recurrence (e.g. the probability that a patient's disease will return after primary treatment); sensitivity and specificity (performance characteristics of the diagnostic test); costs (of both the test itself as well as those which are incurred during the clinical trajectory); and health state utility values (e.g. weighted measures of disease burden, such as the quality-adjusted life year or QALY, which combines both quantity and quality of

[6] Once it was determined that a trial of breast cancer tumor markers was the optimal path to pursue, project organizers eventually went on to conduct EVSI calculations in designing that trial. A discussion of this latter process is beyond the scope of the present paper, but see Thariani et al. (2013) for an account of this process from the perspective of the project organizers.

life in a single measure where QALY= time x utility) (Pettitt et al. 2016). As is typical with VOI modeling in general, the data underlying these parameters was culled from previous clinical trials and meta-analyses, although certain models also included a number of assumptions and expert opinion(s) about those variables for which previous data was unavailable. Uncertainty around those variables was then characterized probabilistically using computer simulations based on prior distributions.

4.3 The dual ontologies of VOI

As a form of *topicalizing* possible avenues for clinical research investment, EVPI calculations are discussed in the VOI literature as being a 'hypothetical construct' (Bennette et al. 2016, Appendix A; Roth 2012:52) -- or what the project organizer quoted above described to me in terms of their being 'useful' but 'not precise.' This latter distinction refers us back to Lampland's (2010) work on formalizing practices, which I have briefly discussed in this paper's introduction. In particular, she finds that numbers in formalizing practices can be 'useful' even while they may lack precision or accuracy, and specifies two ways in which this can be so. For instance, her analysis of the introduction of agricultural accounting practices in Stalinist Hungary reveals a profusion of *false numbers*, where 'the primary task is to learn how to deploy numbers' (ibid.:378). Of course, as time went on, the state eventually imposed harsh punishments for 'systematic misrepresentation' and 'systematic misappropriation' of farm earnings (ibid.:393); but at least during the earliest days of modernization in that country, significantly less emphasis was placed on the 'relative accuracy of the numerical sign' so long as bookkeepers showed that they were 'master[ing] the logic of formal procedures' of accounting (ibid.:378). 'False numbers,' she concludes, 'were tolerated because they were *useful*' for establishing good bookkeeping practices amongst agricultural workers (ibid.:393, italics mine).

A second such type of number Lampland has identified in formalization practices are what she calls *provisional numbers*. Looking specifically to scientific modeling, she finds:

> No matter how much models may be tweaked and refined over time -- as the exigencies of the material and social world are accommodated through experiment and analysis -- models still remain idealizations of empirical conditions. *They are never accurate and thorough depictions of physical or social dynamics.* In fact, the conditional status of models is a general feature of scientific practice. [...] Simplifying conditions and granting unrealistic

> assumptions are part and parcel of the modeling process. This is as true of physics as it is of economics (ibid.:385-6).

Provisional numbers are operationalized as numbers 'used in planning and strategizing' in order 'to assist groups in setting the parameters for tasks at hand and debating their relative merit,' oftentimes 'parad[ing] as stable and fixed indicators, though their provisional status is well known by those responsible for making them' (ibid.:378). One of the central differences between provisional and false numbers is that the former 'are temporary from the outset' while the status of the latter 'only emerges in practice' since determinations of falsity tend to depend 'on the judgment and discretion of auditors' (ibid.:392). In CANCERGEN, VOI calculations had something of a double ontology (Mol 2002), existing at once as both provisional and false numbers.

Consider the description of VOI found in the aforementioned backgrounder document, where it is posited as '[a]n emerging field in economics' that, '[w]ith proper adaptation, we believe [...] will aid researchers in selecting and designing studies so that the health benefit for the information gained is maximized for the research investment' (Carlson et al. 2013, Appendix B). In one way, then, the introduction of VOI approximated Lampland's description of provisional numbers insofar as assisting the selection of parameters and facilitating debate. These calculations simply aimed to furnish stakeholders with quantitative outputs that could help them in apprehending the relative value of different trial options according to a more formal -- and less subjective -- set of criteria. But from another vantage point, the introduction of VOI was the object of an *economic experiment* (e.g. Muniesa & Callon 2007), where project organizers were positioned to test both their ability to adapt VOI for this particular audience of stakeholders, as well as their belief that those outputs would impact how stakeholders viewed the relative value of the tests in question. A central driver of the VOI intervention in CANCERGEN was project organizers' interest in testing whether the additional 'clinical-slash-economic' outputs of these models changed the way the stakeholders viewed the relative value of the diagnostic tests they were tasked with prioritizing. Here, the calculations surfaced at certain other moments as false numbers, subjected as they were to the audits of the stakeholders themselves who were the primary site for the VOI intervention.

For instance, in one situation that occurred during the second in-person stakeholder meeting, a member of the stakeholder group representing the practicing clinician perspective called out the VOI models for being what he termed a bunch of 'statistical malarky.' A project organizer recounted

this event in an interview I conducted shortly after the aforementioned meeting, and she explained to me that she interpreted this commentary as basically asking back to the team of project organizers developing the models: 'Why are we wasting our time with this!?' A second instance of audit surfaced during an interview with a stakeholder who had been invited to represent perspective of a state healthcare payer. Here, she recounted how value of information analysis results had changed the way that some of the stakeholders ultimately ranked their top three priority tests, with breast cancer tumor marker testing coming to occupy a higher ranking than it had been ranked in previous rounds of deliberations, prior to the introduction of VOI. When I asked if she counted herself among that group who elevated the breast cancer tests, she replied that her own ranking remained unaffected by VOI:

> I still had a lot of questions about - well, you had to use assumptions to do the value of information analysis [...] And so I - I had this uneasiness [...] that we would be answering questions about the [breast cancer] *test* [...] but not answering questions about ultimate *health outcomes*. And that's a more important consideration [...] I wasn't sure that - in terms of the ability to help change a health outcome [...] - for me, that test was at the top, even though the value of information showed that. I had a really hard time connecting it. And it could just be that it was above my statistical ability to comprehend [...] I'm the lawyer, not the accountant, and certainly not the epidemiologist or the statistical whiz. But I just had an uneasiness that - said generically - that we were focusing on the wrong thing.

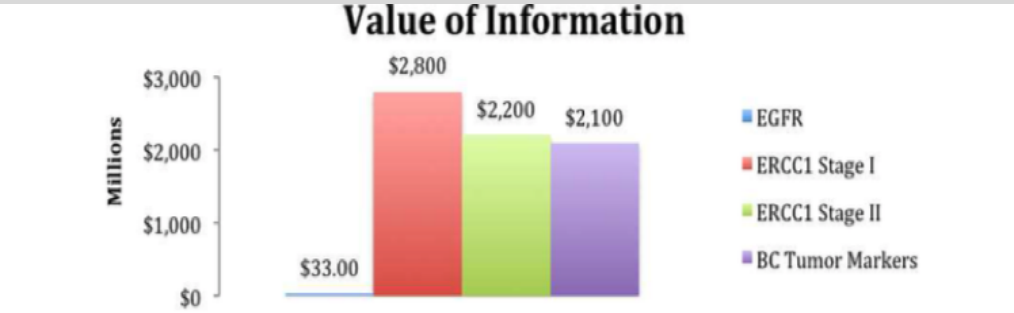
This reference to the stakeholder as the *lawyer* and not the *accountant* subtly references a third normative decision that project organizers made in informationalizing VOI: the decision to display outputs in a monetized information format. Recall that VOI is an approach that values the 'societal pay-off of research,' where, if the value of information derived from further research is projected to exceed the actual costs of investing in and conducting the research, then it is deemed a worthwhile societal investment. In general, the more a society is willing to pay for a year of perfect health, the higher the opportunity costs will be for making the wrong decision; in turn, the expected value of perfect information also increases in proportion to an increased willingness to pay. For CANCERGEN project organizers, one of the advantages of VOI modeling is that

> a variety of willingness to pay thresholds to achieve a given unit of health benefit (e.g. an extra year of life or quality-adjusted life year) can be explored

through this approach, which are key parameters for informing decisions to conduct additional research (Carlson et al. 2013, Appendix B:15).

This is then monetized by multiplying the health gains expected to be produced by better informed future decision making -- in this case, a quality adjusted life year -- by a willingness-to-pay level, which project organizers stated was $150,000, a kind of 'rule of thumb' conventional estimation of what the typical American taxpayer is willing to pay for a year of perfect health in the setting of cancer.

Figure 5.1 VOI of four competing testing strategies to trial
From (Carlson et al. 2013)



Cancer Genomic Application	Affected Population[a]	Probability of Making the Wrong Decision about Testing, %	Consequences of Making the Wrong Decision about Testing	Value of Information[b] (Millions)
EGFR mutation testing in maintenance treatment in advance NSCLC	170,253	12	$1600	$33
ERCC1 testing in early stage NSCLC: stage I	234,051	26	$47,300	$2800
ERCC1 testing in early stage NSCLC: stage II	234,051	42	$22,500	$2200
BC tumor marker testing	416,746	43	$11,700	$2100

BC, breast cancer; EGFR, epidermal growth factor receptor; ERCC1, excision repair cross-complementation group 1; NSCLC, non–small cell lung cancer.
[a]Over 10 years.
[b]Calculated at $150,000 per quality-adjusted life year willingness to pay, discounted at 3%.

In the final analysis, once the VOI results were calculated, project organizers created a bar graph that was circulated to the stakeholders and subsequently published in a peer-reviewed academic journal (ibid.:468; see Figure 5.1). The pecuniary value displayed atop each bar in Figure 5.1 is a function of the health benefits produced by using the intervention under perfect decision-making conditions (that is, with perfect certainty about its use) among the entire population of patients

affected by the given condition (that is, all NSCLC patients for the first, all stage I NSCLC for the second, all stage II NSCLC for the third, and all breast cancer patients for the last), the whole endeavor amounting to a 10-year time horizon.

5. Calculating conventions

5.1 Producing a 'bottom line' value

One might rightfully pause here to pose the question of how, from the heterogeneous grouping of parameters that function as inputs into VOI modeling, its ultimate output is in the shape of a single summary pecuniary value for each strategy under consideration and how this figured into project organizers' normative decisions about calculating the value of information. This was in fact a concern even for the project organizers themselves, as was recounted to me during an interview:

> I think people felt like, when it ends up with a dollar amount -- the value of this information from a trial -- that's something you could think about. But there were a lot of questions [from stakeholders] -- 'Well, really, how did you get to that dollar amount? What?' And that's a huge leap [...]

Here we return to the interview excerpted in the introduction of this paper, where my respondent seemed to liken value of information analysis with cost-effectiveness analysis (CEA), another technique regularly used by health economists to measure the benefits of adopting healthcare interventions. In drawing a parallel between those two techniques, she nevertheless also pointed out to me that CEA outputs are often in quality-adjusted life years (QALYs, described below) while VOI's outputs are in dollars. I questioned why VOI outputs would not simply be presented as QALYs, to which she referred me to the team of project organizers who were actually running the models, saying only that 'they think this is the preferred way to do it.'

Taking her lead, I subsequently raised the question in an interview another project organizer, a practicing clinician and health outcomes researcher who also holds a doctorate in health economics. He pointed out to me that VOI and cost-effectiveness analysis are frequently confused and that this was one of the primary challenges in imparting the logic of VOI onto the stakeholders, since cost-effectiveness 'wasn't the core concept we were looking at.' When I reflected back to him that it seemed to me that there was a certain element of cost-effectiveness inherent to VOI analysis, he replied:

> Oh yeah, it's part of VOI -- although, to be honest, most of the folks in SWOG and most clinicians are somewhat antithetical to using cost effectiveness to make decisions. So we had to be very careful of how we presented the economics information because we didn't want them to feel that we were valuing the study based on cost effectiveness. And we weren't.

Our discussion segued to the palatability of such an economized information format in this particular context, and I asked whether there had been some reflection within SWOG about this novel application of VOI to research prioritization in the US. He went on to say that he understood that

> researchers like VOI because it gives a monetary value to the work they do, which is the way, for better or for worse, people think about anything they do in our society. So, does it add value? Research is a very abstract thing to convey [...] but if you can put a dollar value on research, it makes it much more concrete. Now, I don't *know* this, because I haven't talked to the researchers in SWOG, but I think they are attracted to this notion that we can actually show, in a monetary way, the value -- the payoff -- from the research they do in the clinical trials group.

This belief that monetization is integral to human cognition is further suggested in a published manuscript detailing the elements of CANCERGEN's approach to research prioritization (Carlson et al. 2013). The authors, led by members of the modeling team, note that project organizers not only 'chose a relatively simple application of VOI […] to make the exercise tractable and acceptable to decision makers,' but that stakeholders were also mostly 'amenable' to the introduction of VOI 'in part, we believe, because it was presented simply and across a broad range of topics' -- but also, and more importantly for the present argument, because this simplified application 'presents a single dollar value "bottom line" that provides high-level benchmarking of the maximum potential benefit of further research on divergent technologies and diseases' (ibid:469).[7]

5.2 Conventions of quantification: WTP and QALYs

Yet just as I have shown in the previous section that VOI can be calculated in more than one way -- with EVPI and EVSI being but two instances within a wider methodological approach -- so too does there exist multiple conventions of quantification (Desrosières 1990, cf. Centemeri 2012) that

[7] Such a characterization of human action defined as 'an essentially economizing behaviour' is typical of what Çalışkan and Callon identify as the formalist approach to economics, and is closely linked with neoclassical economics (2009:373).

can be used in formatting the outputs of these models. In other words, there is nothing singular about the 'single dollar value "bottom line"' referred to in the previous quote, a point which the previously quoted project organizer suggested during that same interview:

> [T]he tricky bit is that you have to monetize health to come up with the value, so that means you have to understand what people would be willing to pay for an additional year of life, or [a] quality adjusted year. So we kind of stepped gently around that issue, but that is integral in the evaluation.

In this reply, we find two key concepts -- the quality adjusted life year (QALY) and the willingness to pay (WTP) -- that are necessary for understanding how economization functions within the VOI assemblage as it was deployed in CANCERGEN. But the reply also reflects a tension, albeit a subtle one, between the two because the argument that 'you have to monetize health to come up with the value' is actually contravened by the presence of QALYs in the calculative apparatus. To grasp this subtlety, I will now explain QALY and WTP in turn.

In his 'biography' of the quality-adjusted life year, the sociologist Tiago Moreira (2012, Ch. 3) offers this rather simple definition:

> the QALY is a technique for measuring the benefit obtained from medical interventions by giving a different "weight" on time in different health states. In this, a year of life expectancy in perfect health is worth 1, whereas a year of less than perfect health is worth less than 1. It is argued that QALYs provide a form of currency to assess the extent of the benefits gained from health care interventions, not only in terms of survival but more importantly in terms of the "quality" of the time gained as a result of those interventions (2012:63-4).

The QALY was initially developed during the 1970s in response to a perceived "'failure of the market" to ascertain the value of health care' (Moreira and Maldonado 2017:15). Here, there was a push for methods that sought to elicit preferences (referred to as *utilities)* of individuals 'about health states against which the effects of specific medical interventions could be measured' (ibid.). A prominent example, the Health Utilities Index,

> likened the elicitation of preferences to the choices made by ideal-rational war commanders deciding between two courses of action under uncertainty and imperfect information. [...] One of the key consequences of this was that the QALY was uniquely suited to organizations governed by "command and control" principles, where decision makers could purchase on behalf of

> consumers or users in view of the aggregated health utilities accrued by particular health interventions' (ibid.).

By the early-1990s QALYs had become institutionalized as a dominant metric in health economics. In this latter period, they saw extensive work to standardize their contexts of use and methodological procedures after revealing that despite their numerical format, QALY valuations 'were explicitly normative[ly] value-based' (ibid.). For our purposes, QALYs are notable in that (1) they function as a standardized metric that renders the impact of a given health intervention on the quality of life-years gained through its use; (2) as a utility score, they have a numerical format that ranges from 0 to 1; and (3) they have been used to help decision-makers prioritize investments, such as in purchasing or reimbursing the use of particular medical interventions.

Willingness to pay (WTP) is another convention of quantification in healthcare evaluation (e.g. Grosse 2008; Neumann et al. 2014) with a distinct history unique from the QALY's own emergence -- one that stretches well beyond health-specific fields of implementation. In Donaldson's (2001) review of methods for eliciting patients' WTP in healthcare, he traces the notion's origins to Jules Dupuit, a French engineer who sought a way to measure the social value of projects like the bridge he planned to build across the River Seine in Paris: 'the only real utility,' he concluded, 'is that which people are willing to pay for' (Dupuit 1969 quoted in ibid.:181). Donaldson goes on to say that it was the *maximum* WTP that was most important for Dupuit, which 'represents the price (i.e. money extracted from the consumer) which would be just enough to return him/her to his/her original level of utility. Logically therefore maximum WTP represents the value of the good' (ibid.).[8]

One group of methods for ascertaining willingness to pay is called *stated preferences* -- also known as *contingent valuation* -- where individuals are administered a survey that asks them questions about how much they would be willing to pay for a given good or service (Brown 2003; Fourcade 2011a, 2011b). A second set of WTP methods is referred to as *revealed preferences* (Samuelson 1948) and

[8] Hood (2017) locates the emergence of WTP in US amidst a 40-year period of government debates over policy investments and (de)regulation. Here she finds the search for 'a scientifically sound way to measure the economic value of human life' as policymakers sought to weigh 'how much human lives are worth compared to the cost of saving or prolonging them' - through either investing in particular regulatory policies or by deregulating existing spheres (2017:2). In this analysis, she notes a shift away from a more macroeconomic approach to valuing human life that 'see[s] economic value as the result of tangible contributions to societal wealth' and instead finds value defined 'in terms of consumer sovereignty' where there is economic value of goods 'only insofar as people are willing to pay for them' (2017:12).

assumes that actual purchasing habits are reflective of WTP. In the particular case of healthcare, WTP using stated preferences can be gauged for example by consulting patients or physicians, while structured analyses of payment rates (e.g. Braithwaite et al. 2008) or national budget analyses (Neumann et al. 2015) are indicative of revealed preference methods for gauging a society's WTP. Moreover, Neumann and colleagues (ibid.:796) argue that a society's WTP reflects its cost-effectiveness threshold and that such thresholds 'are used as rough guides to help determine whether particular investments constitute reasonable value' (ibid.). Those thresholds, in turn, are often different from one society to the next, and can even differ from one intervention or indication to the next in a single society -- that is, WTP may be higher in the setting of cancer as compared diabetes, meaning that society on the whole is willing to pay more for health gains in the former than the latter.

5.3 Decision rules and conversion tools

As we have just seen, QALY and WTP are two distinct conventions of quantification for calculating value in healthcare settings -- the former a utility score between 0 and 1, and the latter reflected in actual monetary terms. Historically, however, there has been a de rigueur convergence of these two approaches given that 'QALYs fully become useful [...] only when they are combined with the costs of providing the interventions, from which cost-utility ratios result' (Moreira 2012:64). In this process of making QALYs useful, the resulting cost-utility ratios are typically compared to a WTP to gauge whether the intervention in question presents a good societal investment. In some jurisdictions, such as the United Kingdom, there has in principle been a fairly strict application of a £35,000 WTP to a QALY, so that interventions that cost more than £35,000 to produce a single quality adjusted life year run the risk of not being covered by the the National Health Service, the UK's single-payer national health insurance schema. In this sense, a cost-per-QALY calculation informed by a strict societal WTP functions as a decision rule for adoption and coverage of medical interventions, which 'assumes decision-makers have a fixed value they are willing to pay for a unit of health outcome or that there is a well-established shadow price of health improvements that reflects the opportunity cost of investments' (Grosse 2008:165-66).

The more decentralized healthcare system in the United States, on the other hand, makes such strict decision rules difficult to universally implement and assess since private healthcare providers

have their own proprietary decision-making strategies, while the US federal government explicitly prohibits the Centers for Medicare and Medicaid Services from applying cost-per-QALY thresholds in its provisioning and coverage of publicly funded healthcare. There does exist, however, a lengthy history of using a $50,000-per-QALY threshold in health economics analyse. But as Grosse goes on to point out, this is often approached less as a decision rule and more as a 'rule of thumb' that is said to be 'commonly accepted,' 'commonly cited,' or 'established practice' -- but one that '[m]any observers argue [...] is not based on economic theory, does not derive from a formal expert consensus and is of questionable empirical validity' (ibid.:166). Moreover, there are many other cost-per-QALY thresholds applied in economic analyses in the US, with $100,000 being the second most popular according to Grosse's 2008 review of the health economics literature. Subsequent analyses have found cost-per-QALY thresholds in the US range from $50,000 to $400,000.

In light of this foregoing discussion, and notwithstanding jurisdictional differences and variable strictness of applying cost-per-QALY as a decision rule, the central point here is that QALY and WTP are wholly distinct entities. There are in fact two places that such economic functions can figure into value of information analysis, as is suggested in the following excerpt culled from a published manuscript jointly authored by several project organizers explaining VOI methods as they were deployed in CANCERGEN:

> We used the models to calculate the net benefit for a given treatment strategy based on an estimation of the societal willingness to pay for health gains (e.g., quality-adjusted life year, QALY), which allows for the conversion of the estimated clinical impact of a given strategy into a monetary value. The net benefit of a given strategy was calculated by subtracting the cost of the strategy from the monetized health gains to provide the net benefit (QALY X willingness to pay - cost) (Carlson et al. 2013:465).

In the first place, a cost-per-QALY decision rule is baked into the model's calculations in rendering what is called the *net monetary benefit* of the intervention (cf. Bennette et al. 2016:644-5). Here, the value of information is partially a function of an intervention's adoption only if it achieves cost-effectiveness at a given monetary threshold -- explained above in terms of a societal WTP. But this description is telling in that it also speaks of a 'conversion' of *clinical* impact into *monetary* value -- something quite different from the application of a decision rule as such. So, returning to the previous interview excerpt where the CANCERGEN PI says of value of information analysis that

'you have to monetize to get the value,' what he is really saying is that you have to monetize to get to *monetary* value. Without applying a WTP to a quality adjusted life year, VOI is not unable to produce the value of a healthcare intervention as it relates to research investment, but rather produces a quantitative -- albeit non-monetized -- measure of value, here in the form of a utility score between 0 and 1. It is only by applying a utility score by a given WTP that the outputs of VOI take shape as a monetized information format.

6. Provisional economization

Despite the 'curious resilience' of the $50,000/QALY threshold (Neumann et al. 2015), project organizers leading the VOI component of CANCERGEN performed their calculations using several different WTP levels that were in fact much higher than $50,000. But they also converged upon a single WTP of $150,000 for the purposes of sharing the VOI results back with the stakeholders. On the one hand, this was chosen as the de facto WTP as it was thought to '[approximate] the revealed willingness-to-pay threshold in the United States' although others have argued that the WTP in the setting can in fact be much higher than even $150,000 (Carlson et al. 2013:466; cf. Braithwaite et al. 2008).[9] But on the other hand, they also chose this level for more pragmatic reasons. As one of the project organizers working on the modeling explained to me:

> The problem is that [the] stakeholder group [is] already trying to get up to speed on just what value of information is. There's a lot of stuff that we try to make easier for them to get that. And because willingness to pay can be a tough concept, we chose to tell them about it but if to say, 'Okay, we're just looking at one willingness to pay here.' So they didn't have to consider that as being a major issue and being that it doesn't substantially -- I mean, it affects the magnitude but not too great a deal and doesn't affect the relative ranking. That wasn't where we wanted to spend our energy in terms of educating them.

In other words, picking and sticking with a single WTP was seen as an organizationally efficient means for maximizing tractability of VOI for the stakeholders who, we will recall, had little training in the fundamentals of VOI prior to their recruitment as participants in CANCERGEN.

[9] In fact, VOI was consistently calculated at multiple WTP levels but was restricted to a single value when presenting VOI results to stakeholders to avoid confusion (Carlson et al. 2013:466).

Yet there is another critical element in this passage that speaks to the ontological status and the particular enactment of economic numbers in CANCERGEN. When the respondent speaks of the WTP as affecting the *magnitude* but not the *relative ranking*, he is pointing to the normative decision on behalf of project organizers to present VOI calculations that used only a single WTP across the three clinical scenarios when presenting that information back to the stakeholders -- despite the fact that his team had calculated VOI at multiple different thresholds. He offered a more straightforward explanation in what follows:

> For the purpose of our project, we chose a $150,000 [willingness to pay] but the VOI, this is mostly -- we want to be consistent between models. So this a comparative exercise. The actual value we chose isn't as important for trying to prioritize three different options. We're just saying, 'Okay, health gains are valued the same in each model.' So we wanted to be consistent there. It could have been [a willingness to pay of] $50[,000] but it will still give you one study that's worth more and one study that's worth the least, right? And it impacts the models the same. [...] We looked at a couple of different ones and the rankings don't really change with that multiple thing. So if you did it at $100,000, the results of our models don't change. They don't, like, *categorically* change ranking [...]

I followed up by clarifying if what he was implying was that there was essentially still an order of magnitude difference in value between the different clinical scenarios -- meaning that one was in the millions of dollars and others were in billions of dollars. He continued:

> Right. The actual magnitude of each changes but they don't change places - in which one is the most, second, third in terms of therapy value. The billions don't become millions. It becomes 2.1 [billion at one WTP] versus 1.5 [billion at another WTP] [...] We can put all three on the same graph. You can compare their willingness to pay.

Maintaining consistency of approach, and maintaining differences in orders of magnitude in particular, was a concern that seemed to be shared by several of the project organizers working on running the VOI calculations. Consider the following excerpt from an interview with another of the project organizers who was involved in producing the VOI models. Here, he began recounting how he understood the VOI outputs as impacting the stakeholders' own understanding of the relative value of the three options.

> Before, the stakeholders were kind of like, 'Well, why do you want to study this similar [breast cancer] tumor marker thing that doesn't work?' But when we showed them: 'Look, there'd be a lot of value to knowing [...] you're either consuming costs here, or you're missing helping out women if it works.' So I think that helped them understand that there was greater value.

We will recall that very early on in the project, project organizers presented the stakeholders with a list of six possible technologies that were potential avenues for investing public funds in conducting a large scale trial. One of those trials would be to test a set of breast cancer tumor markers, which at first was ranked very low during earlier rounds of voting but progressively came to take on more importance among the stakeholders over subsequent rounds of deliberation and ranking. Project organizers credited VOI with helping the stakeholders understand the value of the breast cancer tumor markers in a way that did not surface during their previous deliberations, which were guided by more subjective concerns and monographic forms of evidence (e.g. Desrosières 2010). He then moved on to discuss another technology being considered, a test that analyzes the expression of a gene called EGFR in patients with lung cancer, where the presence or absence of a mutation of that gene serves as an indicator of response to particular types of therapy:

> [F]or the EGFR stuff, it's pretty cool and people were excited about it. But the truth is, the data out there is pretty solid: that, clearly, the people with the mutation respond better to a [tyrosine kinase inhibitor, a particular type of drug that intervenes on a cancer-causing molecular pathway]. So [...] do you really need a randomized trial there? And I think the VOR told us, no, it's not as valuable by an order of magnitude.

He summed up the discussion:

> So the ERCC1 [another genetic test for use in lung cancer] and the breast cancer tumor marker, those were -- and this is again an upper bound, not for some specific trial, but for a huge trial - those were like a billion dollars. And the EGFR was -- I don't remember, tens of millions? So it was an order of magnitude difference.

But in a way similar to how the previous interviewees characterized these orders of magnitude differences, so to did this respondent say that, in his view:

> it's not about whether it's $1.7 billion or $1.2 billion or whatever -- it's really kind of order of magnitude effects. Just to make people stop for a minute and go, 'Wow, okay, why is that? Okay, I kind of see that that makes sense.'

> And perhaps that is what caused them to change their ranking. Actually that's what they said -- it affected their rankings.

In sum, what we have seen in this section are several reflections of project organizers who were involved in producing VOI models in CANCERGEN. Their narratives emphasize the importance of an economized information format as one that is presumed to be legible, as well as highlighting the conventional use of WTP levels in a way that deemphasizes any inherent correspondence between the economized information developed and deployed within the local setting on the one hand, and some ontologically stable category of *money* that exists in the 'real world' beyond CANCERGEN. What matters, then, is the ability to construct and relay stabilized differences across multiple competing avenues for research investment. This in turn highlights the conventional, non-market nature of these economic numbers thusly produced in a process I call *provisional economization.*

The provisionality of this economization process becomes all the more evident when comparing how VOI was conducted in the CANCERGEN setting with a subsequent project carried out several years later that transported the method into SWOG itself, which as we will recall was envisioned as the primary site for conducting whatever trial(s) would result from CANCERGEN's innovative prioritization framework. What is particularly interesting in this second instantiation of VOI is that, like CANCERGEN, it entailed a 'stakeholder engagement' element with many of the same features as CANCERGEN's own. In this latter case, the stakeholder group was composed of clinicians, patient advocates, clinical trialists, and statisticians who were involved with the different disease-specific panels in SWOG which are tasked with vetting incoming trial proposals before passing them off to the NCI for final funding approval decisions. Not unlike the ESAG, so too was this latter stakeholder group given access to VOI training sessions and written materials covering the application and interpretation of VOI techniques; invited to comment on inputs and assumptions of the models; and encouraged to ask any clarification questions about the method (Bennette et al. 2016:643).

But in a way quite different from what occurred in CANCERGEN, feedback received from SWOG members during this more recent exercise indicated that stakeholders held conflicting views about building cost-per-QALY decision rules into the calculations, as we have seen was the case in CANCERGEN. As the authors of a manuscript detailing this process recount:

> A slight majority of SWOG members expressed the sentiment that SWOG's mission was to conduct trials that had the greatest potential to improve health regardless of costs. [...] Others felt that drug prices were ultimately out of SWOG's control and were often difficult or impossible to know before a trial started, particularly for new investigational agents. [...] A slightly smaller group of SWOG members [...] felt that costs of many cancer therapies had reached a tipping point and that it was critical for SWOG to start considering how their investments could alleviate or aggravate this problem' (Bennette et al 2016:646).

In response to these comments, the calculations were performed such that 'the decision to adopt a treatment within our simulations [was] made on the basis of health benefits (i.e., QALYs) rather than net monetary benefits [...] In other words, we did not explicitly incorporate a willingness-to-pay threshold in our VOI calculations' (Bennette et al 2016:646-7). Moreover, not only did they not use a net benefit approach, but in presenting the model outputs back to the stakeholders so too was the modeling team moved to 'deconstruct the clinical and economic components of traditional VOI calculations' (ibid.:646). In other words, although there was some cost information -- relating to the clinical scenarios under consideration -- presented to the stakeholders in monetary terms, the actual outputs of the VOI models were not subjected to a post-hoc economization, remaining instead as 'clinical' components. That is, in the form of a QALY; a utility score between 0 and 1.

7. Conclusion: certified worth

It may be argued that *all* economization is to some extent *provisional*. After all, the very term *economization* was developed to emphasize that economic things are always the outcome of a process of active constitution, only ever *provoked* into existence rather than already existing in the world fully formed as such (Çalışkan & Callon 2009; Muniesa 2014). Yet within the wider 'performative turn' in economics (e.g. Callon 2007), far less attention has been paid to how processes of economization relate to qualifications for *market worth*. As we saw in introduction of this paper, the jump from 'economic' to 'market' is perhaps made all too hastily when we are confronted with situations in which things we hold dear -- whether that be the pristine Alaskan wilderness or biomedical research that aims to improve patient outcomes -- are rendered in economic terms.

So the question arises: Does the introduction of value of information analysis in CANCERGEN signals an attempt on behalf of project organizers to import market logics into the prioritization of

publicly-funded cancer clinical research? While at first glance this may seem to be so, it is my contention that this in fact not the case. 'We err,' writes Lampland, 'if we assume that [economic] numbers perform the same task and refer to the same element at all moments' in formalizing practices (2010:379). That is, while formalization practices 'may rely on [economic] numbers,' they nevertheless 'stipulate neither their meaning nor their use a priori' (ibid.). In closely scrutinizing the 'range of pedestrian activities' that constituted the deployment of value of information analysis in CANCERGEN, I have sought to detail the series of normative decisions that rendered VOI tractable to the stakeholders participating in the project, and to highlight the consistent recourse to conventional forms -- including quality-adjusted life years (QALYs) and willingness to pay (WTP) thresholds -- that served as the foundations for what project organizers understood as being a more *formalized* prioritization process.

A second question thus follows from the first: If in the case of economization I have described above is in fact not coextensive with market worth, what kind of qualification(s) *does* it imply? My answer here is that value of information analysis constitutes a process of *governing through standards* which works to *certify the worth of competing options* in a locally circumscribed planning exercise (Thévenot 2014, 2015). Interestingly, even here the literature has tended to directly link certification with markets: Busch, for instance, describes certification as 'a form of trust [...] well suited for a world consisting largely of strangers acting in the *marketplace*' (2011:215, italics mine), while Thévenot speaks of the 'encapsulation of a plurality of engagements' in standard products as

> a requirement of market coordination which assumes that goods (merchandise) and their quality are common knowledge. As a consequence, market worth is placed in a superordinate position with regard to other forms of worth which are reduced to qualities of market goods [...] The market convention of coordination integrates a plurality of other conventions at a lower level in the reified reduced form of objective qualities (2011:60).

But Thévenot also goes on to caution that governing through standards is not simply 'market-worthy and market-driven' but that it is also linked to a 'liberal grammar' in which commonality is constructed as 'a plurality of individuals who, as stake-holders, express their differences in the format of preferred options [...] Then the issue is not market competition through prices, but individual choice which also requires options to be accessible' in a standardized format

(ibid.). In this way, such a mode of governing resolves to a 'coordination infrastructure' in its own right (Thévenot 2015:215).

As a form of governing through standards irreducible to market logics alone, might we understand operations of reduction wholly absent from VOI? Again, I would venture a negative answer here. However, rather than reducing a myriad of forms of worth to a market qualification, as suggested above, what we find in VOI is a reduction to the objectives of *planning*, a particular way of engaging -- linked to the liberal grammar -- in which '[t]he preliminary consolidation of options goes through the formatting of a range of plans -- stakes -- which are the basis of individualized expression and negotiation' (Thévenot 2011:62). In this process, the environment is set up such that the concerns of actors -- like the stakeholders participating in CANCERGEN -- are oriented by and towards the efficient achievement of a formal plan of action: selecting a single clinical trial to be carried out in the Cooperative Group Program. Here, VOI calculations reify options as quantified, economized benchmarks.

In one way, then, VOI reduces 'the appeal to, and debate on, the upper level of the plurality of orders of worth' (ibid.:62). We find this, for instance, in the case of the clinician stakeholder rejecting VOI as a bunch of 'statistical malarkey,' and the lawyer stakeholder who discussed how the modeling exercise seemed to privilege concerns about the technical characteristics of the breast cancer tests over broader considerations of health outcomes, and how she was uneasy about the possibility of focusing on the wrong things (ibid.). In another way, VOI also reduces 'openness to the lower level of familiar engagement and attachments,' such as how each of the stakeholders interact with biomedical knowledge production practices in their daily routines and habits (ibid.). Presumably the concerns of a lawyer charged with making policy decisions are often times quite different from those of a patient living with cancer -- but the two perspectives are brought in line with each other as both parties become equals as 'stakeholders' in this formal prioritization process.

More to the point, understanding VOI analysis as a process of *certifying worth* of biomedical research investment options also points to the conventional and polysemic nature of its calculative apparatus, which presupposes neither universal rejection nor acceptance as a condition for certification (Rosental 2003:640). So while project organizers reported that a 'majority [...] of ESAG members found the VOI analysis to be useful,' only about 'half [...] stated they changed their priority

rankings specifically based on the VOI results' -- which was still sufficient to demonstrate that using 'VOI analyses to inform research prioritization may be a good investment for research organizations and funding agencies given the relatively small cost of such analyses compared with the cost of research studies' (Carlson et al. 2013:468-9). In this way, the exercise worked to certify not only the value of different options to trial but function as a platform for demonstrating the value of VOI analysis in its own right. This precipitated moving the technique further into SWOG itself, where calculations were performed in an altogether different way that, unlike in CANCERGEN, used neither a net benefit framework nor did it economize health outcomes (Bennette et al. 2016; Carlson et al. 2018). Even in this latter case, project organizers recognized that VOI models 'may be overly simplistic representations of complex clinical processes' yet they nevertheless vow to continue assessing in a 'prospective evaluation' whether this process can 'improve SWOG's decision making' (Bennette et al. 2016:650).

In sum, the deployment of VOI in CANCERGEN surfaces as a mode of provisional economization that works to certify the value of options in a locally circumscribed planning exercise. As such, the functional nature of these numbers is indicative of a formalized calculative practice in which we find a reduction to the objectives of achieving a planned course of action. As I have shown in this paper, this is altogether distinct from a reduction to market logics, as some may otherwise suggest given the centrality of economization in VOI as it was enacted within CANCERGEN. Moreover, confining the numbers to the immediate situation of planning also means that they fail to achieve any other public qualification (Boltanski & Thévenot 2006). This is of course not to say that VOI outputs are not capable of *becoming* matters of public concern, as previous studies have shown that investment decisions based upon formalized metrics and made within such confined situations may well circulate outwith their immediate settings and become a site of public critique and contestation (e.g. Moreira 2012). Bearing this in mind, a sensitivity to the multiple forms of reduction that different formalization practices entail enables analyses of valuation practices to stick closer to the situation under consideration which in turn renders a more faithful untangling of these reductions and their political ramifications -- which actors in the situation are capable of addressing in their own critiques of the calculative practices with which they are faced.

8. References

Bennette, C. S., Veenstra, D. L., Basu, A., Baker, L. H., Ramsey, S. D., & Carlson, J. J. (2016). Development and Evaluation of an Approach to Using Value of Information Analyses for Real-Time Prioritization Decisions Within SWOG, a Large Cancer Clinical Trials Cooperative Group. *Medical Decision Making*, *36*(5), 641–651.

Boltanski, L., & Thévenot, L. (2006). *On Justification: Economies of Worth*. Princeton: Princeton University Press.

Braithwaite, R. S., Meltzer, D. O., King, J. T., Leslie, D., & Roberts, M. S. (2008). What does the value of modern medicine say about the $50,000 per quality-adjusted life-year decision rule? *Medical Care*, *46*(4), 349–356.

Brown, T. C. (2003). Introduction to Stated Preference Methods. In P. A. Champ, K. J. Boyle, & T. C. Brown (Eds.), *A Primer on Nonmarket Valuation* (Vol. 99–110). Dordrecht: Springer Netherlands.

Busch, L. (2011). *Standards: Recipes for Reality*. Cambridge, MA: MIT Press.

Çalışkan, K., & Callon, M. (2010). Economization, part 2: a research programme for the study of markets. *Economy and Society*, *39*(1), 1–32.

Callon, M. (2008). What does it mean to say that economics is performative? In D. MacKenzie, F. Muniesa, & L. Siu (Eds.), *Do Economists Make Markets? On the Performativity of Economics*. Princeton: Princeton University Press.

Carlson, J. J., Thariani, R., Roth, J., Gralow, J., Henry, N. L., Esmail, L., … Veenstra, D. L. (2013). Value-of-information analysis within a stakeholder-driven research prioritization process in a US setting: an application in cancer genomics. *Medical Decision Making*, *33*(4), 463–471.

Centemeri, L. (2015). Reframing problems of incommensurability in environmental conflicts through pragmatic sociology: From value pluralism to the plurality of modes of engagement with the environment. *Environmental Values*, *24*(3), 299–320.

Claxton, K. P., & Sculpher, M. J. (2006). Using value of information analysis to prioritise health research: some lessons from recent UK experience. *PharmacoEconomics*, *24*(11), 1055–1068.

Daston, L. (1995). The Moral Economy of Science. *Osiris*, *10*(1), 2–24.

Donaldson, C. (2001). Eliciting patients' values by use of 'willingness to pay': letting the theory drive the method. *Health Expectations*, *4*(3), 180–188.

Dussauge, I. (2015). Valuation machines: Economies of desire/pleasure in contemporary neuroscience. In I. Dussauge, C.-F. Helgesson, & F. Lee (Eds.), *Value Practices in the Life Sciences and Medicine* (pp. 247–264). Oxford: Oxford University Press.

Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a Social Process. *Annual Review of Sociology*, *24*(1), 313–343.

Fourcade, M. (2011a). Cents and sensibility: Economic valuation and the nature of "nature." *American Journal of Sociology*, *116*(6), 1721–1777.

------. (2011b). Price and Prejudice: On Economics, and the Enchantment/Disenchantment of Nature. In J. Beckert & P. Aspers (Eds.), *The Worth of Goods. Valuation and Pricing in the Economy*. Oxford: Oxford University Press.

Grosse, S. D. (2008). Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold. *Expert Review of Pharmacoeconomics & Outcomes Research*, *8*(2), 165–178.

Helgesson, C.-F., & Muniesa, F. (2013). For what it's worth: An introduction to valuation studies. *Valuation Studies*, *1*(1), 1–10.

Hoffman, A. S., Cambrosio, A., & Battista, R. (2016). Comparative Effectiveness Research in Health Technology Assessment. In A. Levy & B. Sobolev (Eds.), *Comparative Effectiveness Research in Health Services* (pp. 57–93). New York: Springer.

Hood, K. (2017). The science of value: Economic expertise and the valuation of human life in US federal regulatory agencies. *Social Studies of Science*, *47*(4), 441–465.

Keating, P., & Cambrosio, A. (2012). *Cancer on Trial: Oncology as a New Style of Practice*. Chicago: The University of Chicago Press.

Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Sociology*, *38*(1), 201.

Lampland, M. (2010). False numbers as formalizing practices. *Social Studies of Science*, *40*(3), 377–404.

Latour, B., & Callon, M. (1997). *'"Thou shall not calculate!" Or how to symmtricalize gift and capital'*. (Unpublished translation of '« Tu ne calculeras pas » ou comment symétriser le don et le capital.' In A. Caillé (Ed.), *Comment peut-on être anticapitaliste?*, La Découverte, La revue du MAUSS #9).

Marks, H. M. (2000). *The progress of experiment: science and therapeutic reform in the United States, 1900-1990*. Cambridge, UK: Cambridge University Press.

Mol, A. (2002). *The body multiple: Ontology in medical practice*. Durham: Duke University Press.

Moreira, T. (2012). *The Transformation of Contemporary Health Care: The Market, the Laboratory, and the Forum*. New York: Routledge.

Moreira, T., & Maldonado, O. J. (2017, October 5-7). Metrics in Global Health. Presented at *Governing by numbers: Key indicators and the politics of expectations*, Martin-Luther-Universität, Halle-Wittenberg, DE.

Muniesa, F. (2014). *The Provoked Economy: Economic Reality and the Performative Turn*. London: Routledge.

Muniesa, F., & Callon, M. (2007). Economic experiments and the construction of markets. In F. M. & L. S. Donald MacKenzie (Ed.), *Do Economists Make Markets? On the Performativity of Economics* (pp. 163–189). Princeton: Princeton University Press.

Nass, S. J., Moses, H. L., & Mendelsohn, J. (Eds.). (2010). *A national cancer clinical trials system for the 21st century: reinvigorating the NCI Cooperative Group Program*. Washington, DC: National Academies Press.

NCI [National Cancer Institute]. (2009a). *ARRA Research and Research Infrastructure Grand Opportunities: Comparative Effectiveness Research in Genomic and Personalized Medicine (NCI RFA-OD-09-004; RC2 Grant)*.

------. (2009b). *Center for Comparative Effectiveness Research in Cancer* Genomics - *CANCERGEN (Project #1RC2CA148570-01)*. Retrieved from https://projectreporter.nih.gov/project_info_description.cfm?projectnumber=1RC2CA148570-01

Neumann, P. J., Cohen, J. T., & Weinstein, M. C. (2014). Updating cost-effectiveness—the curious resilience of the $50,000-per-QALY threshold. *New England Journal of Medicine*, *371*(9), 796–797.

Neyland, D. (2008). *Organizational Ethnography*. London: SAGE.

Pettitt, D., Raza, S., Naughton, B., Roscoe, A., Ramakrishnan, R., Ali, A., … Brindley, D. (2016). The Limitations of QALY: A Literature Review. *Journal of Stem Cell Research & Therapy*, *6*(4), 1–7.

Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton: Princeton University Press.

Rosental, C. (2003). Certifying Knowledge: The Sociology of a Logical Theorem in Artificial Intelligence. *American Sociological Review*, *68*(4), 623–644.

Roth, J. (2012). *Comparative Effectiveness Approaches to Evaluate Pharmacogenomic Technology*. Doctoral Thesis, Department of Pharmacy, University of Washington, Seattle.

Samuelson, P. A. (1948). Consumption Theory in Terms of Revealed Preference. *Economica*, *15*(60), 243–253.

Stark, D. (2009). *The sense of dissonance: Accounts of worth in economic life*. Princeton: Princeton University Press.

Stark, D. (2011). What's Valuable? In J. Beckert & P. Aspers (Eds.), *The Worth of Goods. Valuation and Pricing in the Economy.* (p. 319). Oxford: Oxford University Press.

Thariani, R., Wong, W., Carlson, J. J., Garrison, L., Ramsey, S., Deverka, P. A., … Baker, L. H. (2012). Prioritization in comparative effectiveness research: the CANCERGEN experience in cancer genomics. *Medical Care*, *50*(5), 388–393.

Thévenot, L. (1984). Rules and implements: investment in forms. *Social Science Information*, *23*(1), 1–45.

------. (1995). L'action en plan. *Sociologie Du Travail*, *37*(3), 411–434.

------. (2011). Conventions for Measuring and Questioning Policies. The Case of 50 Years of Policy Evaluations through a Statistical Survey. *Historical Social Research / Historische Sozialforschung*, *36*(4), 192–217.

------. (2012). At-testing, Pro-testing, Con-testing: New perspective on politics, oppression and critique raised by environmental concern. Presented at the *To Engage or Disobey?*, Centre d'Étude Français sur la Chine Contemporaine & Institute of Sociology, Taipei.

------. (2015). Certifying the World: Power Infrastructures and Practices in Economies of Conventional Forms. In P. Aspers & N. Dodd (Eds.), *Re-Imagining Economic Sociology* (pp. 195–216). Oxford: Oxford University Press.

Timmermans, S., & Berg, M. (2003). *The gold standard: The challenge of evidence-based medicine and standardization in health care*. Philadelphia: Temple University Press.

Tuffaha, H. W., Gordon, L. G., & Scuffham, P. A. (2013). Value of information analysis in oncology: the value of evidence and evidence of value. *Journal of Oncology Practice*, *10*(2), e55–e62.

Tuffaha, H. W., Gordon, L. G., & Scuffham, P. A. (2014). Value of information analysis in healthcare: a review of principles and applications. *Journal of Medical Economics*, *17*(6), 377–383.

Wong, W. B., Ramsey, S. D., Barlow, W. E., Garrison, L. P., & Veenstra, D. L. (2012). The value of comparative effectiveness research: projected return on investment of the RxPONDER trial (SWOG S1007). *Contemporary Clinical Trials*, *33*(6), 1117–1123.

Interlude 3

Chapters 4 and 5 grapple with themes of participation and economization respectively, each being a core element of CANCERGEN's prioritization dispositif. Having outlined these activities and framed them in a way that accords with Thévenot's notion of engaging in planned action, we now turn our attention to look at how both of these features were ultimately combined in the work of prioritizing a single technology with which to proceed with in a subsequent phase of designing a CER study.

As we argue in the following chapter, although quantification may reduce the relevant realities to single benchmarks -- as is the case with value of information analysis -- even ostensibly more 'democratic' forms of (e)valuation may operate at the level of a similarly reduced set of realities. In this way, while we as social scientists may be inclined to valorize participatory modes of multi-criteria decision making, the latter are still not coextensive with the work of making things public.

Rather, by approaching situations from the perspective of the evaluative good at stake, a certain parity emerges between qualitative and quantitative modes of valuation in such a way that firmly implants them both within a coordinative regime of engaging in plans. This also stands to evidence the heterogeneity found within a given regime of engagement -- here, the planning regime -- a point that has heretofore remained undertheorized in Thévenot's own work, and is perhaps generalizable to other regimes of engagement as well.

Chapter 6

Which technology to trial? The evaluative politics of research prioritization in oncology

Abstract

This paper takes up the question of what is worth knowing in biomedical research. It focuses on a multi-institution initiative called CANCERGEN whose mission was to develop a platform for streamlining the prioritization and design of large-scale clinical studies in oncology -- combining a deliberative stakeholder engagement process together with quantitative-economic modeling practices. Zooming in on one phase of this project, and drawing on both published and internally circulated CANCERGEN documents, a succession of discrete *evaluative moments* are identified during which project participants sought to answer the question of what counted as a valuable technology -- that is, a technology worth investing scarce public funds in order to learn more about its clinical usefulness. In tracing CANCERGEN's syncretic approach to (e)valuation, the analysis draws on Thévenot's regimes of engagement framework to contrast a *politics of elicitation* with a *politics of performation* that combined to valorize a set of breast cancer diagnostics for proceeding with the design of a clinical trial. Despite both existing within an ostensibly singular regime of planning, they nevertheless signal a certain heterogeneity internal to the regime itself, evidenced by the quite distinct affordances and capacities for coordination and valorization that contrasting modes of quantitative and deliberative evaluation offer.

1. Introduction

Which road to follow? Such is the question Laurent Thévenot poses in the title of a 2002 book chapter, wherein he examines a proposal to build a new highway and tunnel connecting France with Spain through the Pyrenees Mountains. His exploration, he says, is 'all about what counts, or should count, as a "good road" and what is the reality of such a road' (8). In asking this question, Thévenot adds roads to a long list of other phenomena that are exemplary of a renewed emphasis within the social sciences over the past few years for studying practices of valuation and evaluation in contemporary society (Cefaï et al. 2015; Lamont 2012; Mol 2013). Despite the diversity of empirical sites that have attracted students of (e)valuation, and an equally wide range of analytic perspectives

that have been used to approach these empirical issues, one could say that there is a broad and concerted effort to approach the value question from a more pragmatic, practice-based perspective (Hutter & Stark 2015).

This paper seeks to elaborate upon current discussions on the production and enactment of value(s) while forging a closer link between valuation studies and the sociology of quantification (e.g. Centemeri 2012; Espeland & Stevens 2008). To do so, I focus on valuation in the domains of healthcare and biomedicine (e.g. Dussauge et al. 2015; Will & Moreira 2010) and, more specifically, the prioritization and design of clinical trials in oncology. Some recent scholarship has, in fact, taken up the theme of valuation in studying biomedical research design. Helgesson and colleagues (2016), for example, have studied valuations of experimental designs in 'traditional' randomized clinical trials (RCTs) as compared with more recent innovations in designing biomarker-informed proteomics trials (BTTs). Their approach 'examin[es] trends in the design of biomedical experiments and the intertwining, balancing, and hierarchisation of different values and measures of value for biomedical research' and 'attempts to shine light on shifting yardsticks, values, and pressures in the contemporary research landscape' (ibid.:157). Yet before clinical trial *designs* can be decided, a host of first-order decisions must first be made. Primary among these questions is deciding which medical interventions should be studied in the first place. Perhaps unsurprisingly, we witness a range of competing valuations and (modes of) evaluation that factor in to providing possible answers to this question, as well.

Accordingly, the present analysis examines the role that numbering practices play in the development of clinical research studies; their syncretization with other more qualitative and deliberative modes of evaluation; and how these combine to inform (e)valuations about prioritizing technologies to trial. As such, the case presented here thus takes up Thévenot's at once philosophical and empirical question of 'which road to follow' in quite a different context, shifting the site of analysis from a transportation project in France to a clinical research project in contemporary American biomedicine. Here, I describe the work of the Center for Comparative Effectiveness Research in Cancer Genomics (CANCERGEN), a three-year research initiative funded by the US National Cancer Institute in 2009. The project was a multi-institution collaboration that sought to execute 'a consensus process with multiple stakeholders [to] develop a comprehensive evaluation,

assessment and design process to prioritize emerging cancer genomics technologies' for conducting clinical trials (Deverka et al. 2012a:361).

This endeavor was quite a complex undertaking in terms of both its organizational structure and the diversity of deliverables the project pursued. It involved the collaboration of project members from four different institutions (an academic cancer hospital and research center; a university-based health economics and outcomes research department; a nonprofit organization specializing in stakeholder engagement exercises; and a large national cancer clinical trials research network) as well as a group of stakeholders from a diverse set of backgrounds. It also entailed several different phases during its existence, and each phase entailed its own set of (e)valuations: in the early work of identifying and convening the stakeholder group, there was the question of who should count as a 'stakeholder,' while in the final step -- in a way similar to the processes that Helgesson et al. (2016) describe -- there were the valuations of the actual designs of a large-scale randomized controlled trial.

Given CANCERGEN's complexity, this paper zooms in on one phase of the project that began following the selection of stakeholders to sit on the External Stakeholder Advisory Group (ESAG) and concluded with the selection of a single technology for which project members would design an RCT. In this sense, very much in line with Thévenot's account of the Somport tunnel project, the CANCERGEN project members and stakeholders were reflexively asking a very similar question: Which (genomic) technology to trial? In recounting this process, I draw from project-related documents (including meeting summaries and technical documents produced by various project members for internal use) as well as official publications authored by project participants in order to home in on three distinct features of this middle phase of CANCERGEN:

1. A series of *evaluative moments* during the CANCERGEN prioritization process wherein project members and stakeholders sought answers to the very questions of what counts -- or should count -- as a good technology to trial. The first of these moments began with the early work of conducting a landscape analysis to identify all relevant candidate technologies; this was followed by the work of composing a manageable list of candidate technologies, which were then the subject of several subsequent rounds of stakeholder voting and ranking;

the process concluded with the moment the decision was made about which technology to move forward with for conducting a large-scale clinical trial.

2. The set of *evaluative equipments* that actors drew on in justifying their claims for why particular technologies should or shouldn't be examined in the context of a large-scale clinical trial. At first, this was a situation of deliberative exchange among and between project members and stakeholders, where discussions were informed in equal measure by stakeholders' own knowledge and judgment derived from their experience as decision-makers in the healthcare field as they were by the descriptive charts and relevant data about the candidate technologies that project members provided to the stakeholders as background reference. However, in time there was a marked shift towards numerical practices and quantitative modes of reasoning. This took the form of the production of econometric models; efforts to educate stakeholders about the rationale, methods, and outcomes of these models; and attempts to study the impact these models had on the way that stakeholders evaluated and ranked the top three candidate technologies.

3. The (genomic) *technologies* themselves as sites where different evaluative equipments produced different 'matters of concern' during the successive unfolding of these evaluative moments (Latour 2004). My approach here accords with recent work in valuation studies, which impels us to 'go beyond the mere observation of a plurality of valuation processes and to actively deal with their interrelations,' and asks: 'When are the outcomes of different valuations competing and when do they lead to substitutions? […] When are they unrelated and when can they co-exist in a space without any mutuality apart from temporal and spatial co-presence' (Kjellberg et al. 2013: 22). In tracing the path of one particular set of technologies -- a group of three biomarkers that are used to indicate a possible recurrence of breast cancer in patients once they have been declared cured -- I seek to show the how the deployment of a plurality of (e)valuation processes aided their movement from being of (statistically) least importance for stakeholders to becoming a prime contender for subsequent efforts at designing a large-scale clinical trial, all the while displacing other

candidate technologies that stakeholders had ranked much higher during earlier phases of the prioritization process.

This (co)existence of multiple evaluative equipments within the succession of evaluative moments that made up the CANCERGEN prioritization process refers us back to the pragmatic approaches to (e)valuation discussed in the opening paragraph, characterized by their attentiveness to specific elements of valuation. These elements have distinct consequences for the way that value is produced, and it is to this theoretical discussion that we first turn our attention.

2. (E)valuations of a lesser scope: a tale of two pragmatisms

Simply mentioning pragmatic approaches to valuation and evaluation signals a distinct orientation towards these phenomena. There are in fact at least two different streams that label themselves as such. On the one hand, there is scholarship on (e)valuation in the American pragmatist tradition. This body of work draws its influence from John Dewey and especially his 'flank movement,' which 'makes the distance between value and its measure collapse in an analytically constructive manner' and thus calls for 'a shift in subject matter from value (or values) to valuation, considered explicitly as an action' (Muniesa 2012:25). In other words, there is no singular 'value', nor plural 'values', nor the distinction between 'economic value' and 'social values' as the so-called 'Parson's Pact' would have it (e.g. Stark 2009:7). Instead, there are only things, situations, events, etc. that are said to be valuable, and whose value hinges on the processes and practices of *valuation*. Similarly, a pragmatist approach to *evaluation* considers it to be a 'practical judgment' that is 'characterized by a particular type of object: it bears on things to do or to be done in relation to a given situation' (De Munck & Zimmermann 2015:114). From the perspective of American pragmatism, and especially Dewey's pragmatic approach to valuation, the central distinction to be made here is that *valuation* is understood an attempt to identify what is good, whereas *evaluation* 'is a matter of justification and proof, for it involves judgment likely to be publicly defended' (ibid.:122).

French pragmatic sociology has also taken up the themes of valuation and evaluation, most notably in Thévenot's collaborative work with Luc Boltanski on orders of worth (e.g. 1999; 2000; 2006). Here, the two authors come quite close to the American pragmatist approach to evaluation in terms of their sensitivity to the interactional and deliberative quality of evaluation and the role that

public justification plays therein. According to this perspective, certain forms of engagements -- where 'engagement' is understood as both the cognitive processes actors use to grasp their environment and the actors' pledging in the search for a guaranteed good -- are open to a very public form of critique. In this most conventional *regime of justification*, actors routinely draw from a limited number of *orders of worth* in evaluating situations and making critical claims, with each order based on a foundational text of political philosophy (Thévenot 2010). Actors thus increase the strength and generality of what may at first seem to be very situated claims by relating those claims to more universal conceptions of the 'common good.' Yet as Thévenot reminds us, not all engagements operate at such a generalized and public level; and so sociological analyses of evaluation must thus also be sensitive to more limited forms of coordination and their operational dynamics.

It is with this claim that Thévenot departs from his work with Boltanski and builds a new conceptual model that he calls pragmatic *regimes of engagement* (henceforth: ROE). In his discussion of the Somport tunnel project, asking which road to follow leads to an exploration of the different regimes of engagement that produce the value of a road. In that first and most publicly legible regime of *justification*, regimes are qualified according to orders of worth: the highway is said to be valuable from the *market* order of worth because it encourages competition and free markets in Europe by improving transportation; but it also can be said to be worthy from an *industrial* order because it upgrades the road and is seen as an investment that 'builds for the future, making planning possible because it will function reliably'; and still other orders can be brought to bear on the valuation of a road (Thévenot 2002:64). And yet the ROE framework is attuned to (e)valuations that operate at a far more localized scale than those found within the most legitimate forms of qualification just discussed. The key here is 'to take the notion of agency beyond the regime of justification […] and focus on the different ways in which human beings *engage with* their environments of artifice or nature' (Thévenot 2002:68).

Thévenot thus introduces two regimes that operate beneath the level of public justification. The second regime he adds to the model is the *regime of familiar engagement*, which is the most intimate regime. Here, a road is not said to be valuable because of its relation to a specific order of worth, and in fact it may not a road at all, but rather takes the form of a *path*: 'Neither designed nor planned as a

functional instrument, the path emerges as a nonintentional result of acquaintance of human beings with a milieu of human and nonhuman beings. The path is created through habitual frequenting as much as physical topography. Indeed it may simply reflect a pattern of wandering' (ibid.:70). The third regime is the *regime of regular planned action* -- also called the *regime of engagement in a plan*, or simply the *normal regime*. In this regime, a road's value comes neither from a specific order of worth, nor from individuals' relationships with their dwelled-in environments, but rather from the fact that it functions 'as a tool for going from one place to another. That is the end of the story.' (ibid.:73).

As can be seen from this brief summary, the ROE framework is prepared to deal with (e)valuations that operate at a far more localized scale than the types of evaluations found within the most legitimate forms of qualification discussed above. Of particular importance for this model is that it accounts not only for the discursive elements of engagements, but also the spatial *environments*, *material equipments*, and *agencies* that produce and are produced within the coordinative processes that constitute (e)valuations. These elements are also tied to specific temporal horizons of each of the three regimes of engagement: the regime of justification has a variable temporal horizon, depending on the kinds of work being invoked, while the familiar regime is turned to the past and the regime of engagement in a plan deals with projections into the future (Thévenot 2013).

American pragmatist perspectives on evaluation are similarly concerned with a parallel set of phenomena. For Dewey, a significant element of evaluation practices is their 'relational dimension' that 'link[s] *environmental* conditions to *ends and means*, *persons* and *things*' and this leads to the conclusion that evaluative judgments are made 'through the critical exercise of ordering the findings of inquiry' (ibid.:125). A second element of evaluative judgments is their temporal dimension, which brings the influence of past occurrences and possibilities for future outcomes to bear on the present situated action. Whence another commonality between Thévenot's ROE framework and Dewey's work on valuation: a focus on the interactions between individuals and their environment in (e)valuation practices.

Both American pragmatism and French pragmatic sociology are also attuned to the deliberative character of evaluations, which I have already briefly touched upon. For Dewey, deliberation is the 'weighing of various alternative desires (and hence end-values) in terms of the conditions that are the means of their execution, and which, as means, determine the consequences actually arrived at'

(quoted in de Munck & Zimmermann 2015:125). When there are a plurality of values in play -- for example, when economic, creative, or social values can be used to define 'good work' in workplace evaluations -- there is a requirement for a

> space for deliberation in which the different stakeholders […] can voice what quality […] means to them, in terms of ends and means. This internal space for deliberation should, furthermore, exist within a broader space that includes external stakeholders […] Only by strengthening and improving situated democratic debate […] can we hope to progress in the process of reconciling [a plurality of competing] values (ibid.:131).

On the other hand, Thévenot has outlined the plurality of values in terms of orders of worth; in the regime of justification, deliberation occurs between individuals who mobilize behind different orders of worth in waging public critiques, while coordination takes the form of a compromise between multiple (and often) competing orders by making them compatible. For example, returning to the Somport tunnel project example, local actors introduced a domestic order of worth to the market order that was the central mode of justification for the project. Within this market-domestic compromise, the road is not a 'truck corridor', as the centralized government in Brussels argued for, but is proposed by local actors to afford access to local tourist sites and for local trade, 'favor[ing] a road going to and ending within the valley: a way of entering rather than passing through it' (Thévenot 2002:65).

Yet there are also at least two critical distinctions between Dewey's and Thévenot's conceptualizations of deliberation. The first is that Dewey views deliberation as an inherent good, whereas Thévenot's ROE framework opens up the space to probe for different forms of deliberation. In the regime of justification, deliberation is aimed at a common good, while in the regime of engagement in a plan, the scope of deliberation is much more limited. Engagement here takes the form of 'strategic engagement', where the negotiation of interests excludes other possible modes of participation and opens itself to new forms of oppression (Cheyns 2010:4). Relating to this first difference, Dewey's approach leads to his conclusion that including 'external stakeholders' in deliberation is also a universally positive element of its functioning, thereby improving the likelihood of reconciling competing evaluations. Conversely, for Thévenot, the agency produced within the regime of justification is said to take the shape of the 'moral subject' acting in the common good,

while in the familiar regime the agent is a 'personality' that is distributed throughout their milieu and strives for a good that 'has to do with taking good *care* of this *accommodation*' (Thévenot 2002: 65). Alternatively, a 'stakeholder' is not simply any 'actor,' but is a very specific type of actor that is engaged within the regime of engagement in a plan, and whose agency obtains in their 'voluntary capacity, individual intention, choice, the project, interests and strategy to achieve certain goals and objectives' (Cheyns 2010:8). Accordingly, Thévenot's regimes of engagement framework occurs as a perhaps better suited to accounting for the nuances that exist in various forms of engagement where competing (e)valuations are deliberated in differentially equipped environments.

3. Prioritizing candidate technologies: deliberating uncertainty

As we have just seen, pragmatic approaches to valuation and evaluation conceptualize these phenomena as modes of action and interaction, achieved through deliberative processes and the production of certain forms of agency. That stakeholder engagement was one of the main forms of novelty that CANCERGEN claimed to introduce into the work of prioritization and assessment in clinical trial design points precisely to this particular feature of evaluation. More specifically, this emphasis on stakeholder engagement accords with Thévenot's typification of the actors and agencies produced within the *regime of engagement in a plan*. Yet at the very outset of CANCERGEN, a very important evaluative decision needed to be made: what qualifies someone to be a 'stakeholder'? While an in-depth discussion of this issue is beyond the scope of this paper, it should be noted that CANCERGEN's stakeholder recruitment process was based on a framework developed by members of the project, leading to the identification of five core decision-making constituencies within the healthcare system. The External Stakeholder Advisory Group was ultimately comprised of 13 members: three healthcare payers; three practicing clinicians; two policy-makers; two patients/consumers; one regulator; and two representatives from the pharmaceutical and diagnostic industries (Deverka et al. 2012b:2).

3.1 Making a list: the first evaluative moment

Once the stakeholder group was convened, it was time to get down to work. If the overarching goal of CANCERGEN was to conduct a clinical trial on a genomic technology, and the stakeholders were to be the drivers of this decision-making process, then making a manageable list from which

they could select a final candidate was of utmost importance. Prior to initiating the formal stakeholder engagement part of the project, members first developed a formalized prioritization process called a landscape analysis and used this method to identify all possible technologies on the research and clinical horizon that could be subjected to rigorous investigation by conducting a large-scale clinical trial. What ultimately began as a list of over 4000 studies identified in a search of the Internet publication database Medline was systematically reduced through several rounds of exclusion criteria to a shortlist six candidate tests culled from the top five most prevalent forms of cancer (Thariani et al. 2012b, Table 6.1).

Table 6.1 List of six candidate genetic tests identified by landscape analysis
Adapted from (Esmail et al. 2013:117)

Candidate Technology	Purpose
EGFR Mutation Testing for Erlotinib Therapy after 1st Line Chemotherapy in Non-Small Cell Lung Cancer	Disease prognosis and identification of the patients with mutations that are most likely to benefit with erlotinib maintenance therapy
ERCC1 Expression Testing for Platinum-Based Adjuvant Therapy in Resected, Early Stage Non-Small Cell Lung Cancer	Disease prognosis and identification of ERCC1-negative patients who are most likely to benefit from platinum-based adjuvant chemotherapy
BRAF mutation Testing in Colorectal Cancer to Guide Use of Cetuximab and Panitumumab	Disease prognosis and identification of patients with mutations that are good candidates for therapy with cetuximab and panitumumab
EGFR Gene Copy Number (FISH) Testing and Cetuximab Therapy in Advanced Non-Small Cell Lung Cancer	Disease prognosis and identification of patients with high-risk profiles for treatment with
Gene Expression Profiling (GEP) in Multiple Myeloma to identify high risk patients for more aggressive therapy	Disease prognosis and identification of patients with high-risk profiles for treatment with bortezomib
Breast Cancer Tumor Marker testing using CEA, CA 15-3, and CA 27.29 to detect recurrence after primary breast cancer therapy	Prognostic for recurrence of breast cancer and can lead to earlier treatment of recurrence

After the shortlist was finalized, a set of Test Target Profiles (TTPs) was then developed, which are described by project members as 'an easy to read two-column chart listing the criteria by which each test will be measured on one side and descriptive terms on the other' (CANCERGEN 2010a:

3). The TTPs included information on several different priority-setting criteria relating to each test: the population health impact; the adequacy of the current standard of care; issues of analytical and clinical validity; benefits and harms of the test and associated treatment; economic impacts; evidence of need for a randomized controlled trial (RCT); clinical trial implementation and feasibility issues; and market factors (Esmail et al. 2013:118). The purpose of the TTP was not simply to offer background information on the tests under consideration for a clinical trial, but also to have those criteria function as structuring elements within the stakeholder discussions that followed the distribution of the TTPs. In other words, they were meant to serve as resources upon which stakeholders could draw in making discursive claims for why they believed one technology should be carried forward with a clinical trial over other possible technologies.

3.2 Checking it twice: the second evaluative moment

Several weeks prior to the first in-person stakeholder meeting, the TTPs were distributed to the ESAG members along with a brief summary of the TTP's contents; this latter document also included additional details about the tests, information on insurance coverage policies for the tests, and other information that could be useful for the stakeholders. As a follow up, stakeholders were sent an online survey in which they were asked to choose their top three out of the six total candidate technologies that resulted from the landscape analysis and were asked to rank their selections in order of perceived need for conducting a clinical trial (i.e. first, second, and third place). The results of this first round of voting were then tabulated, revealing a three-way tie for the first-place ranking (Table 6.2). As can be see in Table 6.2, the composite means of stakeholder's rankings were also calculated, placing EGFR mutation testing as the highest ranked (CM=3.5) with BRAF (CM=3.4) and ERCC1 (CM=3.3) following in second and third place rankings, respectively. It should be noted that only the top three ranked tests would continue on as the foci of stakeholder discussions in subsequent rounds of the prioritization process. This means that, based on these first-round rankings alone, breast cancer tumor markers would no longer be considered for conducting a clinical trial: not only were they not in the 'shortlist' of the top three candidates, but they also displayed the lowest calculated composite mean ranking (CM=3.2) of all possible choices and were ranked sixth (e.g. last place) by the greatest number of stakeholders (n=4).

The first in-person ESAG meeting was held at the University of Washington in Seattle in June 2010, shortly after this first round of voting took place. During this meeting, CANCERGEN project members facilitated a discussion among stakeholders in order to elicit their reasoning behind the initial rankings they gave. For reasons elaborated below, I will discuss only the top three ranked technologies, and the breast cancer tumor markers that were deemed to be of statistically least importance to stakeholders.

Table 6.2 First round of voting/ranking [Online]
Prior to first in-person ESAG meeting, May 2010
Adapted from (CANCERGEN 2010b: 5)

RANKING >>> **TEST:**	**1st**	**2nd**	**3rd**	**4th**	**5th**	**6th**	**Composite Mean Ranking**
EGFR Mutation Testing	3	2	1	0	2	1	3.5
BRAF Mutation Testing	3	1	3	2	0	0	3.4
ERCC1	3	3	1	2	0	0	3.3
EGFR Fish Testing	0	1	2	3	1	2	3.2
Breast Cancer Tumor Markers	***0***	***1***	***2***	***1***	***1***	***4***	***3.2***
GEP in MM	0	1	0	1	5	2	--

As evidenced by the presence of both EGFR mutation testing and ERCC1 expression testing in the top three ranked technologies, stakeholders expressed a broad interest in genomic tests within lung cancer. Those who favored the EGFR mutation testing -- the highest ranked test overall in terms of composite mean -- as a first choice argued that it was 'potentially an ideal case' for a clinical trial (CANCERGEN 2010b: 7). Stakeholders based their evaluations on the test's ability to identify a small subset of patients with a specific mutation who would benefit from a particular drug, called erlotinib, which targets that mutation in killing the cancer cells (i.e. its population impact); a strong association between the test and the increased survival rates in patients with the mutation (i.e. its clinical validity); its ability to enrich the population of responders to the targeted therapy, erlotinib (i.e. its clinical benefit); and the ability to avoid costly therapy by limiting its use to those patients with the genetic mutation (i.e. its economic impact). Similarly, interest in ERCC1 expression testing was driven by a concern for the population impact of testing -- both in terms of the frequency of genetic variant and the overall prevalence of non-small cell lung cancer -- and was also based on

potential clinical benefit in terms of the test's ability to effectively identify patients with stage I NSCLC who could benefit from specific types of chemotherapy.

The main drawbacks of both EGFR mutation and ERCC1 mutation testing were related to practical issues surrounding clinical trial implementation and feasibility: since only about 10% of patients with non-small cell lung cancer have the EGFR genetic mutation that the test identifies, the total number of possible patients who could enroll in the trial is limited at the outset; the fact that there were already many clinical trials for lung cancer patients more generally only compounded the concern over having adequate numbers of patients for a trial on either of the tests to be able to produce useful information in a timely fashion (Esmail et al. 2013:5). As well, there were specific concerns about the ERCC1 expression test's analytic and clinical validity and so stakeholders expressed that the test may require further validation studies before a large scale clinical study could be conducted.

Those stakeholders who ranked BRAF mutation testing in colorectal cancer as their top candidate drew primarily from the economic impact of the test, as well as ability to prevent harm: as a prognostic indicator, BRAF mutation testing can identify patients who are unlikely to benefit from anti-EGFR antibody therapy, since most patients with the BRAF mutation do not also have a second common genetic mutation (the KRAS gene, for which anti-EGFR antibody therapy is used to treat). BRAF mutation testing can then save patients the costs, side effects, and time spent on antibody treatment that is unlikely to be effective. During the conversation about BRAF testing, one stakeholder representing the payer perspective stated 'that of all six tests, this particular test was generating the greatest interest among his medial policy team because they are fielding so many questions about it from nurses and patients' (CANCERGEN 2010b: 10). It was also thought that clinical trial implementation and feasibility would be limited, although stakeholders who gave BRAF testing a low ranking did so based on its limited population impact in terms of the overall low frequency with which the mutation occurs in colorectal cancer patients.

As we see in Table 6.2, breast cancer tumor markers received the lowest ranking in this round of voting. Yet nearly all of the stakeholders present at the first in-person ESAG meeting were concerned about the fact that the three breast cancer biomarkers under consideration -- CEA, CA15-3 and C 27.29 -- were being routinely used by physicians while monitoring patients for

follow-up, despite the low levels of evidence supporting clinical benefit in this context. Stakeholders 'saw a trial as an opportunity to potentially highlight an ineffective technology' (Esmail et al. 2013: 119). Given the poor performance of some of these tests, there were also concerns about the potential for harm in using them. One clinical study of the CA 15-3 biomarker test that was cited several times during stakeholder deliberations showed that the test had a sensitivity of only 36%, meaning that the test only correctly identified 36% of patients as having evidence of recurrence, while the other 64% of patients with evidence of recurrence went undetected (i.e. a false negative).

One of the project members also mentioned that, despite the poor and/or wholesale absence of evidence about the use of breast cancer biomarker testing in monitoring recurrence, as many as 40% of breast cancer patients will get one or more of the tests over the course of their follow-up: 'Given this reality, and a growing appreciation for the psychological impact of living and interacting with the health care system as a breast cancer survivor, some stakeholders became more interested in studying this test' (CANCERGEN 2010b: 9). Moreover, another project member disputed the economic impact that was listed in the TTP, saying that it was likely much higher than reported in the document due to the fact that physicians will use these tests several times during the course of follow-up for a single patient, compounding the cost issue many times over. Adding to this financial complexity were market concerns, which they defined as a high degree of pressure on payers to cover the use of these tests even in the absence of data supporting their usefulness.

Despite these concerns, there was also a fair amount of discussion among stakeholders and project members about whether CANCERGEN should conduct a clinical trial that would demonstrate that a test *shouldn't* be used in practice or covered by insurance companies. Referring to the lung cancer genetic tests discussed above, the object of conducting a clinical trial of either of these two screening tests would be to show their *clinical utility*: how they contribute to better patient management and improved patient outcomes by more effectively targeting therapy to or away from populations with certain genetic expressions. The question for the breast cancer biomarkers was entirely different, and so was the potential for 'political backlash.' This concern was especially palpable during this discussion in light of the controversy that had erupted -- just a month prior to CANCERGEN being funded -- when the U.S. Preventive Services Task Force reversed its stance on mammogram testing and recommended against routine screening for women between the ages of 40

and 49. This decision was met with an uproar from patient groups and consumer awareness campaigns (ACS 2009). One stakeholder representing the patient advocate perspective stated that, from a clinical trial implementation and feasibility perspective, such a trial may in fact be feasible based on these political concerns. On the other hand, another stakeholder representing the payer perspective stated: 'Realistically, payers can't make change until they have some legitimate study backing, especially when you are talking about things that have already disseminated and have a lot of emotional influence such as this' (Esmail et al. 2013:120). Thus, political concerns such as those expressed by the patient advocate stakeholder were also wrapped up in more technical concerns as well as an appeal to individuals' emotional investments.

Generating evidence that would be of sufficient quality so as to influence such a wholesale move *away* from using the breast cancer biomarker tests was not seen as an easy task: stakeholders believed that a clinical trial of these tests was likely to be very large, very long, and very expensive. That the tests were already being widely used in clinical practice despite the existence of evidence indicating their possibly limited usefulness in monitoring disease recurrence only further complicated matters. Furthermore, the combination of the limited and discouraging evidence about these tests' analytic properties, and wholesale lack of evidence indicating that early detection of breast cancer recurrence would impact treatment decisions or overall survival, led many stakeholders to ask the question: 'Why bother studying something if there is no benefit to patients?' (CANCERGEN 2010b:9). At the end of this discussion session, representatives from the payer, test developer, and clinician stakeholder constituencies 'remain[ed] unconvinced about the scientific merits of pursuing a prospective CER study of these breast [cancer] tumor markers, suggesting instead a range of alternative study designs such as the creation of a biobank, retrospective and survey-based studies.' (ibid.: 10).

Table 6.3 Second Round of voting/ranking, During first in-person ESAG Meeting, June 2010
Adapted from (Esmail et al. 2013: 118)

1st	ERCC1 Expression Testing in NSCLC
2nd	EGFR Mutation Testing for Erlotinib
3rd	BRAF mutation testing in Colorectal Cancer

Following this discussion, and in light of the information that was presented and shared during the two-hour discussion period, stakeholders were asked to once again rank their top three candidate technologies for prioritization (Table 6.3). The rankings were in fact identical to the first round of online voting that occurred prior to the first ESAG meeting, hinting that stakeholders were unswayed by the 'rich discussion' surrounding the breast cancer tumor markers (ibid.:9).

3.3 The third evaluative moment: (re)framing as (re)evaluation

Project members then probed stakeholders in order to better understand the respective rankings in this second round of voting. It was here that a 'lively discussion' played out: clinician and patient advocate stakeholders explained that they favored doing a study on breast cancer tumor markers over BRAF testing because of the possible psychological and public health consequences of testing. Conversely, 'many of the other stakeholders questioned doing a study with tests they felt had little evidence of benefit,' with a stakeholder representing the payer perspective stating that these breast tumor markers 'are a waste of resources'; in some accordance with the clinician and patient advocate stakeholders, however, they also invoked the 'instinctive fear with breast cancer' which perhaps warrants paying closer attention to these tests (CANCERGEN 2010b:11).

The CANCERGEN Principle Investigator then stepped in to offer his own observation that the divergent opinions among the stakeholders about the usefulness of the breast cancer tumor markers mirrored the more general split between clinicians who routinely order and use the tests to monitor recurrence, and genomics experts who profess a more cautious approach. He went on to draw an analogy between breast cancer biomarker testing and the prostate specific antigen test (PSA) test, the latter of which was approved by the US FDA in 1994 and is used to test 30 million American for early detection of prostate cancer annually at an expense of $3 billion. Two studies published in 2009 -- the same year that CANCERGEN was funded -- showed the limited usefulness of the PSA test: one American study reported that PSA screening failed to lower deaths from prostate cancer, while a European study reported a .6% reduction in mortality and concluded that for every one man who was helped by PSA testing, 48 patients received therapy that was deemed unnecessary at best and harmful at worst. Summarizing the results of the European study, Dr. Otis Brawley, the Chief Medical Officer of the American Cancer Society, opined that the PSA test 'is about 50 times more likely to ruin your life than it is to save your life' (Ramsey 2010). Thus, during the preceding

stakeholder discussion, the CANCERGEN PI also professed his favor for conducting a clinical trial on breast cancer tumor markers because, like the PSA test, 'it has been around for a while and it complements the emerging tests on the market' and could even be 'the "PSA" of breast cancer' in its potential public health impact (CANCERGEN 2010b:11).

Table 6.4 Third round of voting/ranking
During first in-person ESAG meeting, June 2010
Adapted from (Esmail et al. 2013: 121)

Rank	Test	Rationale/Prioritization Criteria
1st	ERCC1 Expression Testing in NSCLC	• high population impact • high clinical validity • ability to quickly obtain end points in a trial due to rapid disease progression
2nd	EGFR Mutation Testing for Erlotinib in NSCLC	• relatively high population impact • high clinical validity • large opportunity costs since erlotinib is expensive and has significant harms
3rd	Breast Cancer Tumor Markers	• high population impact • high degree of dissemination into current clinical practice • low clinical validity • opportunity to curb inappropriate use of markers and realize savings

This discursive reframing of the importance for conducting a clinical trial on the breast cancer tumor markers had a significant impact on the way the tests were ranked during the third round of voting: using a show of hands, half of the stakeholders ranked ERCC1 expression testing as their number one candidate technology, while the other half ranked EGFR mutation testing as their first priority, (Table 6.4). However, 'the additional discussion about the merits of breast cancer tumor marker testing resulted in a final higher priority ranking score by the stakeholders for this category of testing versus the BRAF test' (ibid.:12).

Because only eight of the thirteen stakeholders were present at this first in-person ESAG meeting, a fourth round of online voting to prioritize the top three candidate technologies was held two weeks after the meeting, which all ESAG members participated in (Deverka et al. 2012b:189).

This final round of voting confirmed that the top three ranked technologies from the deliberations at the Seattle ESAG meeting were those that most interested the group as a whole.

3.4 Interlude: deliberation as evaluation in CANCERGEN

Approximately one year following the sunsetting of CANCERGEN's funding period, a journal article co-authored by several CANCERGEN project members was published, describing the development and execution of the project's prioritization process (Esmail et al. 2013). In part, the article summarizes many of the stakeholder discussions that we have reviewed in this section, which were focused on the elicitation of stakeholders' respective rationales for ranking the candidate technologies as they did prior to, during, and in the follow-up after the first in-person ESAG meeting in June 2010. Despite the appearance of the BRAF mutations in two different tables within that article -- one highlighting the initial list of six technologies that resulted from the landscape analysis, and the second which lists the shift in rankings between pre-meeting and post-meeting (e.g. breast cancer tumor markers overtaking BRAF as the third ranked candidate technology) -- no mention is made of stakeholder discussions about the importance and possible complications of conducting a clinical trial on that test. That is, rather than focusing on stakeholders' downgrading of the BRAF test, the majority of emphasis in this article is placed on the upgrading of the breast cancer tumor markers which brought it among the top three ranked candidate technologies and into the next round of prioritization.

This accords more generally with the shape that stakeholders' discussions had taken during the ESAG meeting. Recall our brief review of the three top three candidate technologies coming into the ESAG meeting: the two lung cancer tests, ERCC1 and EGFR, and the BRAF test in colorectal cancer. A closer examination reveals that, both in the published literature emanating from the CANCERGEN project as well as the internal documents prepared for project members' and stakeholders' reference (e.g. the CANCERGEN priority setting memo, the Test Target Profiles, and the official summary of the first ESAG meeting), discussions about these technologies dealt primarily with their technical and practical specificities in terms of their relative strengths and weaknesses as candidates for a large scale clinical trial. In other words, an interesting outcome of the multi-criteria list outlined in the TTPs was that it set up a *logic of justification* from which stakeholders were asked to draw in their discussions, and offered its own internal set of *orders of worth* that in many

ways parallels the formatting of arguments that can be made within regime of justification (Thévenot 2001).

In these discussions, many of the criteria by which stakeholders were asked to evaluate the need for a clinical study on each of the technologies came under scrutiny. For example, how should 'population impact' be defined? Does it refer to the overall incidence of the type of cancer that each of the tests corresponds to? Does it more specifically mean the prevalence of the genetic variant that the test actually captures? Or is simply a question of the number needed to treat, understood as the number of patients that need to be tested for one patient to benefit, as compared with a control in a clinical trial? Similarly, does 'potential clinical benefit' point to the ability to choose the most effective therapies for patients? Does it mean minimizing the cost, time, and of life that patients lose when subjected to ineffective treatments? Or does it refer to using tests to reduce harm, for example by using the test to identify a small subset of patients that might benefit from an aggressive treatment with many side effects while sparing the rest of the patient population who don't have that mutation from the treatment's deleterious effects (Esmail et al., 2013).

Notwithstanding these questions, one must also consider a second way these criteria differ from how orders of worth have been conceptualized by Boltanski and Thévenot (2006[1991]): stakeholders' justifications were rarely scaled up to the public level, but rather stayed very much at the level of the project and were directed towards fulfilling the project's ultimate goal of selecting a single technology with which to proceed with in designing and conducting a clinical trial. In sticking very close to fulfilling the project's aims, as Thévenot's *regime of engagement in a plan* would have it, they also had the effect of limiting the range of possible justifications to those most concerned with the *utility* of the tests in terms of their clinical, population, patient, and economic impacts, and the functional/practical concerns as regards the feasibility of designing and conducting a successful trial. And yet, as the engagement process continued, the discussion of the breast cancer tumor markers took on a somewhat different tone.

In addition to the issue of utility, there was a shift in how the tests were conceptualized. There were distinctly political concerns that involved stakeholders and project members collectively envisioning what others in the world outside the CANCERGEN project would need from a trial in order to be confident in their decisions *not* to use the breast cancer tumor markers for monitoring

recurrence of disease, if this is what the trial's outcome was to indicate. This was especially evident in the discursive reframing offered by the Principal Investigator of CANCERGEN during the final moments of the discussion on the breast cancer tests, which was sufficient enough for the stakeholders to move those tests to a third place ranking. This moment is also noteworthy because it was the final point in the CANCERGEN process that stakeholders' own expertise and judgments would be the central referents during the deliberative work of prioritizing candidate technologies. The introduction of quantitative modeling, discussed in the next section, represented a significant shift in the *evaluative equipment* of the decision-making *environment* within the CANCERGEN prioritization process. It is marked by the presentation of a novel *information format* -- where numbers become the *outputs* of a complex mathematical modeling process rather than the *inputs* to stakeholders' evaluations -- and the production of a form of *agency* based on the *performation* of stakeholders' calculative capacities rather than the *elicitation* of their experiential judgments.

4. (E)valuating value: modeling the worth of technologies to trial

As we saw in the previous section, the process of making the final list of three candidate technologies was accomplished through the use of qualitative methods such as those just described, with quantitative practices and numerical representations primarily coming in two somewhat limited forms: either as statistics about the tests' various properties and applications, or as descriptions of stakeholder's rankings of technologies (e.g. as the composite means in Table 6.2). However, this was only one element of a much more complicated model of stakeholder engagement that had a myriad of possible inputs, outputs, methods and processes. The model suggested various methods 'for combining the various inputs to arrive at a decision' and highlighted the multiplicity of *informational formats* that can be deployed in this process: these consist of qualitative methods (e.g. nominal group techniques, facilitated workshops/meetings, and stakeholder decision analysis) as we saw earlier, but also a series of quantitative methods (including questionnaires, Delphi-methods, multi-criteria mapping, and value of information modeling) (Deverka et al 2012b:187).

With regard to this latter method, a publication discussing these aspects of the model states: 'More recently, techniques such as value of information (VOI) analyses have been adapted for use with stakeholder groups to assist with decisions such as priority setting or study design trade-offs, as

a complementary strategy to inform decision-making' (ibid.). At first glance, the mention of VOI methods seems to be simply a passing reference to the emergent use of a novel technique that might be used as an information aid in decision-making. Yet in reality, the introduction of VOI modeling signified a major turning point in the CANCERGEN prioritization project as a whole.

Approximately five months following the first in-person ESAG meeting in June 2010, CANCERGEN investigators began the process of integrating VOI modeling -- also often referred to as 'value of research analysis (henceforth: VOR) -- within CANCERGEN. According to project members, this type of modeling is useful for 'estimat[ing] the benefit of investing in additional studies to determine whether a particular test or treatment should be brought to clinical practice' and they can be helpful for decision-makers who are seeking to 'maximize the impact of research portfolios on medical care and human health. (Deverka et al 2012a:362).

4.1 The value of information: 'A pragmatic yet limited approach'

With roots in statistical decision theory dating back to the 1950s and 1960s, VOI involves the use of methods from decision analysis and economic theory 'to estimate the humanistic and economic value of performing additional research to better understand the safety, efficacy, and cost of technologies and medical interventions' (Carlson et al. 2013: 463). According to Centemeri (2012, 2015), economic techniques of valuation such as VOI can involve compromises between multiple forms of value (such as the 'economic' and 'humanistic' forms of value just mentioned), but they nevertheless produce an 'objective' outcome in the form of a single quantitative value. In VOI, this is a monetized, numerical value representing the amount of uncertainty that can be solved by conducting additional research on the use of a specific test, which, in this fourth evaluative moment within CANCERGEN, involved a choice between ERCC1, EGFR, or breast cancer tumor marker testing.

Introducing VOI methods into the CANCERGEN prioritization process was no easy task. In recounting the development of their models, CANCERGEN project members noted that it

> took 7 researchers (averaging 40% full-time equivalent) approximately 7 months to develop the VOI models, educate the ESAG about VOI, present and discuss the results, and have the ESAG revote as to their top 3 tests. The estimated cost for this effort was approximately $250,000. The bulk of the

> time was spent developing the decision-analytic models and conducting the VOI analyses (Carlson et al. 2013:466).

From this description, it becomes apparent that VOI was a significant 'investment in form' (Thévenot, 1984), not only in terms of financial resources and time spent developing the models, but also with regards to the investments in educating stakeholders about VOI methods, discussing with them the various modeling procedures, rationales, and outcomes of these models, and then having them reevaluate their rankings based on this knowledge and information.

Stakeholders were first sent a 14 page 'model brief' explaining the way the models were built, including background information, the rationales for looking at the specific tests, cost-effectiveness acceptability curves, and model inputs, along with a four page 'Value of Information Analysis Brief,' which included a summary of the general purpose, background and rationale for using VOI techniques; a section explaining the theory behind the techniques; an 'approach' section; and a post-hoc VOI analysis that CANCERGEN project members performed on the National Emphysema Treatment Trial. This example was used as a heuristic tool to help the stakeholders understand what it means for a trial to be worth doing from the perspective of VOI. The Analysis Brief also included the following explanation of what VOI analysis was intended to do:

> For each of the genomic test topics selected by the ESAG, the VOI approach will capture what is already known about their respective clinical and economic impacts, while more fully characterizing the uncertainties surrounding these estimates in the form of a decision problem. This type of analytic framework will 1.) help focus the discussion on those parameters which are most critical for understanding how use of the test affects clinical management of the oncology patient, 2.) promote an informed debate of the relative merits of future research investments to study these issues, and 3.) also help decide which endpoints should be included in a proposed comparative effectiveness study. However, professional judgments will inevitably be required and VOI models should be viewed primarily as a tool to support more explicit decision-making and to use scarce resources more efficiently (Carlson et al. 2013, Appendix B: 17).

Once the stakeholders reviewed these background materials, the CANCERGEN project members who led the modeling efforts held a teleconference with the ESAG members where they reviewed the different models.

4.2 Quantifying uncertainty: the fourth evaluative moment

As briefly mentioned above, producing VOI models for the top three ranked candidate technologies -- EGFR mutation testing and ERCC1 expression testing in lung cancer, and breast cancer tumor marker testing -- rendered the clinical impact of these testing strategies in a numerical, monetary value. However, VOI modeling is quite a complex endeavor: the value of information can be calculated at multiple levels, rendering it conceptually opaque for individuals with no formal background in this area, which is further complicated by the esoteric language used to describe these various levels and processes. For example, at its most general level, VOI is calculated in terms of the Expected Value of Perfect Information (EVPI), which is essentially the value of new information obtained from a trial performed with unlimited subjects, and provides an upper bound on the returns to be gained in conducting further research.

As explained in the Analysis Briefing Paper issued to the ESAG members:

> VPI can be interpreted as the maximum amount we would be willing to spend to learn about the benefits and risks of a particular test or therapy. Although having perfect information is not possible, VPI is the upper level on the value of further research and as such can serve as an initial threshold to support research funding decisions (Carlson et al. 2013, Appendix B: 17).

Beneath this level is the Expected Value of Sample Information (EVSI):

> Whereas the VPI is an upper level on the returns from all possible further research, the value of study information (VSI) represents the expected value of research before conducting a single trial. The expected benefits of research can be compared with the expected costs of carrying out that research, i.e. the costs of sampling. If the cost of new research is less than the expected value of information from the study (EVSI) the trial is worth the expense. The VSI can be regarded as the societal pay-off to research, and can be calculated for a range of samples, sizes and alternative designs (ibid., italics in original).

Referring back to the lung volume reduction surgery trial mentioned above, stakeholders were presented with a short vignette that contextualized these complicated ideas. The EVPI for this study was calculated as being between $31 and $607 billion dollars based on two different willingness-to-pay (WTP) thresholds (e.g. how much an individual is willing to pay for a given unit of

health benefit, frequently calculated in terms of quality-adjusted life years, or QALYs). At those same two WTP levels, the EVSI of the lung volume reduction surgery trial was calculated at $3.5 and $7.3 billion, respectively. Comparing the actual cost of conducting the trial, which was approximately $59 million, one can reasonably conclude that this is much less than the numerical, monetized amount of uncertainty that the trial solved (which was $3.5 billion at the lowest willingness-to-pay threshold), indicating that the trial 'was a very good research investment compared to the cost of covering the surgery' (ibid.:16).

In contrast to this heuristic example, only the EVPI was calculated in the models for each of the top three candidate technologies, a decision that was made by the modeling team 'in part to make the exercise tractable and acceptable to decision makers but also due to resource and time constraints' (Carlson et al. 2013:469). Further attesting to the complexity of VOI, and the need to make the approach 'tractable' for the ESAG members, the CANCERGEN project members in fact calculated the VOI for the three tests at several different WTP levels, but knowing that it would likely render the process of introducing modeling even more difficult, they only used a WTP threshold of $150,000-per-QALY when presenting the results to the ESAG. This latter rate understood as approximating the amount of money a typical American is willing to pay for an additional year of life in perfect health; the net benefit of each testing strategy was then calculated by monetizing the health gain (QALY x WTP) and subtracting the cost of the testing strategy from that amount (ibid.:466).

The model inputs comprised the total affected population in each disease area, as well as available data from previous clinical studies on each of the tests, including information about testing costs, drug costs, other associated medical costs, overall survival and disease progression rates, health state utility measures, and adverse events experienced by patients during treatment. In certain cases, data was not available for specific model parameters: for example, in the breast cancer tumor marker model, there was no data about whether earlier detection led to increased survival times, and so the modeling team 'utilized expert opinion to provide the necessary estimates' (Carlson et al. 2013, Appendix A: 11). Based on the outputs of the VOI modeling (Table 6.5), it is evident that the breast cancer tumor markers had the highest probability of making the wrong decision about testing (43%) out of all three technologies, while both the overall value of information ($2.8 billion) and the

consequences of making the wrong decision about testing ($47,300) were highest for ERCC1 expression testing.While the results were presented during the ESAG teleconference, stakeholders had the chance to raise questions to the modeling team about the models' different facets; these were reported as being primarily technical in nature, and focused on how the modeling team calculated the affected populations and willingness-to-pay thresholds as well as uncertainty around the estimates incorporated into the VOI analyses (Carlson et al. 2013:466-8).

Table 6.5 VOI results of highest-ranked testing strategies
Adapted from (Carlson et al. 2013)

Cancer genomic application:	Affected population	Probability of making wrong decision about testing %	Consequences of making wrong decision about testing	Value of information (millions)
EGFR mutation testing in maintenance treatment of advanced NSCLC	170,253	12	$1600	$33
ERCC1 testing in early stage NSCLC: stage I	234,051	26	*$47,300*	*$2800*
ERCC1 testing in early stage NSCLC: stage II	234,051	42	$22,500	$2200
Breast cancer tumor marker testing	416, 746	*43*	$11,700	$2100

Recall that up until this VOI-focused teleconference, which took place in November 2010, the three candidate technologies were ranked as 1) ERCC1 expression testing in NSCLC, 2) EGFR mutation testing for erlotinib in NSCLC, and 3) breast cancer tumor markers to monitor recurrence. Following the discussion and presentation of VOI results, stakeholders were asked to reflect upon the information presented during the teleconference and to re-rank their top candidate technologies. Here, EGFR testing dropped to a third-place ranking, while the breast cancer tumor markers replaced them in second place (Table 6.6).

Table 6.6 Final ranking, post-VOI analysis *Adapted from (Carlson et al. 2013:468)*	
1st	ERCC1 Expression Testing in NSCLC
2nd	Breast Cancer Tumor Markers
3rd	EGFR Mutation Testing for Erlotinib in NSCLC

Stakeholders reported that their rankings changed based on the 'orders of magnitude difference' between the VOI calculations for EGFR testing ($33 million) versus breast cancer tumor markers ($21 billion), the fact that the probability of making the wrong decision about testing was significantly higher for the breast tumor markers (43%) than for EGFR testing (only 11%), as well as the difference in the total affected populations (170,253 for advanced NSCLC versus 416,746 for breast cancer). The numbers that the VOI models produced were also influential in stakeholders' reevaluating the rankings of candidate technologies because there 'appeared to be substantially more value in studying what they had considered an older, inappropriately used technology than they had realized before being presented the VOI analyses' (Carlson et al. 2013:468). Overall, in a survey given to stakeholders after the presentation of the VOI models, seven of the 13 the ESAG members stated that their rankings changed specifically because of VOI results, while 9 of the thirteen reported finding VOI to be useful (ibid.).

4.3 Second interlude

Numbers can sometimes be curious things. In closely examining the sources of the model inputs for each of the three highest ranked candidate tests, one finds an interesting discrepancy: each of the 39 inputs used for the ERCC1 model and all but one of the 36 inputs for the EGFR models are culled from the clinical literature, with only the rate of hospitalizations for febrile neutropenia -- a side effect of therapy -- being listed as an 'assumption.' Conversely, only 17 of 30 inputs for the breast cancer tumor marker model have source citations, with the 13 remaining inputs being listed as either 'expert opinion,' 'calculation,' or 'assumption' (Carlson et al. 2013, Appendix A).

This is not to say models aren't always already value-laden instruments, but rather that in the specific empirical case I have discussed here, the model that had the single greatest impact on increasing the *value* of a candidate technology in the eyes of the ESAG members was also the *most* 'value-laden' of the three. That is, not only did the model act as a device in producing the value of

the breast cancer tumor markers such that they became the second most valuable technology to trial in *monetary* terms, but it served the simultaneous function of producing the value of the tests in stakeholders' (re)evaluations during the final round of ranking, raising it from third to second in the overall rankings; in the end, it was in fact the most feasible trial to carry forward to the study design phase (Thariani et al. 2013). This model also had a recursive effect insofar as the numerical, economized representations of the value of the breast cancer tests were a site for stakeholders to reevaluate their personal, qualitative evaluations of the tests: recall that what some stakeholders initially considered an 'older, inappropriately used test' was, at the end of the prioritization process, a 'matter of concern' for moving forward with a clinical trial.

5. Conclusion

In this paper, I have explored the admittedly complicated work that goes into answering the question: Which (genomic) technology to trial. Drawing inspiration from pragmatic approaches to (e)valuation in general, and especially Thévenot's conceptualization of regimes of engagement (2001), I have sought to show how, within a single regime -- in this case, the *regime of engagement in a plan* -- multiple (e)valuative practices can exist.

As Thévenot reminds, different 'investments of forms' produce varying 'forms of the probable,' understood as 'different constraints on what can be proved and offered as relevant evidence'; despite any differences in these 'forms of the probable,' they nonetheless 'rest in part on material evidence and the involvement of objects' (Thévenot 2002:56-7). Moreover, he concludes, the different cognitive forms that are produced within these different 'forms of the probable' lead to different modes of coordination based on how they provide access to reality. The reality afforded by VOI models in this final evaluative moment was quite different than the earlier evaluative moments discussed in the previous section. Where reality was, during the first three evaluative moments, apprehended through a deliberative format focused on the *elicitation* of stakeholders' personal experiences and judgments, and further informed by descriptive statistics, reality took on a much more 'durable' format in the shape of objective figures rendered as numerical, economic outputs of opaque, econometric models. Nonetheless, these numbers had a marked influence on producing the value of breast cancer tumor markers: in the process, the role of stakeholders was reconfigured by

way of CANCERGEN project members shifting their focus from elicitation to the *performation* of stakeholders' calculative capacities as they were taught, however quickly, to think like the models (e.g. Callon 2009).

However, despite the deployment of both qualitative and quantitative evaluation practices in the CANCERGEN prioritization process, their existence within the same *regime* does not imply a coexistence in any temporal or even spatial sense, strictly speaking (Kjellberg et al., 2013:22). Rather, what I have called 'evaluative moments' are in fact more akin to *dispositifs* (Foucault, 1977) that see valuations take very different forms in terms of the set of technologies, evaluative and cognitive equipments, environments, and agencies that are constantly shifting during the succession of these moments, and thus implies a shifting of modes of coordination and temporality. In the example I have presented here, VOI models are produced in the present, using data collected in the past, such that the future is rendered more navigable; stakeholders also draw from their own past experiences to negotiate in the present and project themselves into the future -- but sometimes also out into the world, as they seek to move beyond the regime of engagement in a plan to envision what a worthwhile technology looks like, and what the reality of that technology might be, in a more general, justifiable sense.

Nevertheless, these different evaluative moments produce different possibilities for coordination. In the politics of *elicitation*, there exists a certain generative capacity; in addition to the six evaluative criteria that the TTPs specified in earlier phases of prioritization, CANCERGEN project members analyzed stakeholders' discussions and discovered that three additional criteria were produced within these discussions. These new criteria included the importance for tracking *patient-reported outcomes*; difficulties relating to the *recruitment* of patients for publicly-funded clinical trials (as opposed to those sponsored by the pharmaceutical industry); and *clinical trial ethics*, which questions the conditions under which treatment could be withheld in a clinical trial on the basis of a person's genetic profile, and how acceptable would this be to institutional review boards during study start-up (Esmail et al., 2013:120).

In some sense, these new criteria reach both closer to the *regime of familiar engagement*, in terms of attempting to track the lived experience of patients through patient-reported outcomes, and also into the *regime of public justification* in trying to envision how actors external to the CANCERGEN project

would respond their recommendations and trial designs. In contrast, the politics of *performation* vis-à-vis the production of value of information modeling is much more unidirectional and thus has a far more limiting effect on coordinative possibilities. 'Values' are let in the back-door of the models, as it were, in the form of 'calculations' and 'expert input' and 'assumptions,' only to produce a numerical, monetized representation of a candidate technology's value aimed at streamlining the decisions required to achieve the endpoint of selecting a single technology to subject to a large scale randomized controlled trial. As a more objective representation of a technology's value, possibilities for negotiation are much more limited, thereby entrenching actors' coordinative actions firmly within the *regime of engagement in a plan*.

6. References

ACS [American Cancer Society]. (2009, November 16). Press release: American Cancer Society Responds to Changes to USPSTF Mammography Guidelines. Retrieved December 30, 2015, from http://pressroom.cancer.org/index.php?s=43&item=201

Boltanski, L., & Thévenot, L. (1999). The sociology of critical capacity. *European Journal of Social Theory*, *2*(3), 359–377.

------. (2000). The reality of moral expectations: a sociology of situated judgement. *Philosophical Explorations*, *3*(3), 208–231.

------. (2006). *On Justification: Economies of Worth*. Princeton: Princeton University Press.

Callon, M. (2009). Elaborating the notion of performativity. *Le Libellio d'AEGIS*, *5*(1), 18–29.

CANCERGEN [Center for Comparative Effectiveness Research in Cancer Genomics]. (2010b). CANCERGEN Priority-Setting Process.

------. (2010a). Center for Comparative Effectiveness Research in Cancer Genomics Expert Stakeholder Advisory Group Meeting Summary, 14 June 2010.

Carlson, J. J., Thariani, R., Roth, J., Gralow, J., Henry, N. L., Esmail, L., … Veenstra, D. L. (2013). Value-of-information analysis within a stakeholder-driven research prioritization process in a US setting: an application in cancer genomics. *Medical Decision Making*, *33*(4), 463–471.

Cefaï, D., Endreß, M., Nicolae, S., & Zimmermann, B. (2015). Special Issue: Sociology of Valuation and Evaluation. *Human Studies*, *38*(1).

Centemeri, L. (2012). The contribution of the sociology of quantification to a discussion of objectivity in economics. In J. C. Caldas & V. Neves (Eds.), *Facts, Values and Objectivity in Economics* (pp. 110–125). Abingdon: Routledge.

------. (2015). Reframing problems of incommensurability in environmental conflicts through pragmatic sociology: From value pluralism to the plurality of modes of engagement with the environment. *Environmental Values*, *24*(3), 299–320.

Cheyns, E. (2010, 22-24 February). Technical rationality and (de) politicisation of standards. Multi-stakeholder initiatives in sustainable agriculture. Presented at *Governing through Standards - An International Symposium*, The National Museum of Denmark, Copenhagen, DK.

De Munck, J., & Zimmermann, B. (2014). Evaluation as Practical Judgment. *Human Studies*, *38*(1), 113–135.

Deverka, P. A., Lavallee, D. C., Desai, P. J., Armstrong, J., Gorman, M., Hole-Curry, L., … Veenstra, D. L. (2012). Facilitating comparative effectiveness research in cancer genomics: evaluating stakeholder perceptions of the engagement process. *Journal of Comparative Effectiveness Research*, *1*(4), 359–370.

Deverka, P. A., Lavallee, D. C., Desai, P. J., Esmail, L. C., Ramsey, S. D., Veenstra, D. L., & Tunis, S. R. (2012). Stakeholder participation in comparative effectiveness research: defining a framework for effective engagement. *Journal of Comparative Effectiveness Research*, *1*(2), 181–194.

Dussauge, I., Helgesson, C.-F., & Lee, F. (2015). *Value Practices in the Life Sciences and Medicine*. Oxford: Oxford University Press.

Esmail, L. C., Roth, J., Rangarao, S., Carlson, J. J., Thariani, R., Ramsey, S. D., … Deverka, P. (2012). Getting our priorities straight: a novel framework for stakeholder-informed prioritization of cancer genomics research. *Genetics in Medicine*, *15*(2), 115–122.

Espeland, W. N., & Stevens, M. L. (2008). A sociology of quantification. *European Journal of Sociology*, *49*(3), 401–436.

Foucault, M. (1980). Confessions of the flesh. In C. Gordon (Ed.), *Power/knowledge: Selected interviews and other writings 1972-1977*. New York: Pantheon Books.

Helgesson, C.-F., Lee, F., & Lindén, L. (2016). Valuations of experimental designs in proteomic biomarker experiments and traditional randomised controlled trials. *Journal of Cultural Economy*, *9*(2), 157–172.

Heuts, F., & Mol, A. (2013). What is a good tomato? A case of valuing in practice. *Valuation Studies*, *1*(2), 125–146.

Hutter, M., & Stark, D. (2015). Pragmatist Perspectives on Valuation: An Introduction. In A. B. Antal, M. Hutter, & D. Stark (Eds.), *Moments of Valuation: Exploring Sites of Dissonance* (pp. 1–14). Oxford: Oxford University Press.

Kjellberg, H., Mallard, A., Arjaliès, D.-L., Aspers, P., Beljean, S., Bidet, A., … Geiger, S. (2013). Valuation studies? Our collective two cents. *Valuation Studies*, *1*(1), 11–30.

Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, *38*, 201–221.

Latour, B. (2004). Why has critique run out of steam? From matters of fact to matters of concern. *Critical Inquiry*, *30*(2), 225–248.

Muniesa, F. (2012). A flank movement in the understanding of valuation. *The Sociological Review*, *59*(s2), 24–38.

Ramsey, S. D. (2010, 17 May). Transforming Medical Care with Personalized Medicine: Promise and Obstacles. Presented at the International Society for Pharmacoeconomics and Outcomes Research - 15th Annual International Meeting, Atlanta, GA.

Stark, D. (2009). *The sense of dissonance: Accounts of worth in economic life*. Princeton: Princeton University Press.

Thariani, R., Henry, N. L., Ramsey, S. D., Blough, D. K., Barlow, B., Gralow, J. R., & Veenstra, D. L. (2013). Is a comparative clinical trial for breast cancer tumor markers to monitor disease

recurrence warranted? A value of information analysis. *Journal of Comparative Effectiveness Research*, *2*(3), 325–334.

Thariani, R., Wong, W., Carlson, J. J., Garrison, L., Ramsey, S., Deverka, P. A., … Baker, L. H. (2012). Prioritization in comparative effectiveness research: the CANCERGEN experience in cancer genomics. *Medical Care*, *50*(5), 388–393.

Thévenot, L. (1984). Rules and implements: investment in forms. *Social Science Information*, *23*(1), 1–45.

------. (2002). Which road to follow? The moral complexity of an 'equipped' humanity. In J. Law & A. Mol (Eds.), *Complexities: Social Studies of Knowledge Practices* (pp. 53–87). Durham: Duke University Press.

------. (2005). Pragmatic regimes governing the engagement with the world. In T. R. Schatzki, K. Knorr-Cetina, & E. von Savigny (Eds.), *The practice turn in contemporary theory* (pp. 64–82). New York: Routledge.

------. (2007). A science of life together in the world. *European Journal of Social Theory*, *10*(2), 233–244.

------. (2013). The Invested Human Being: Four Extensions of the Notion of Engagement. In M. S. Archer & A. Maccarini (Eds.), *Engaging with the World: Agency, Institutions, Historical Formations* (pp. 162–180). Milton Park: Routledge.

Will, C., & Moreira, T. (Eds.). (2012). *Medical proofs, social experiments: Clinical trials in shifting contexts*. Surrey: Ashgate.

Chapter 7

Conclusion

1. Overview

This thesis has detailed the emergence and enactment of comparative effectiveness research (CER), a novel approach to healthcare evaluation that became a going concern in both the US and internationally circa 2009. At present, it stands as the sole scholarship focused on CER within the fields of medical sociology, science and technology studies (STS), and valuation studies and is the only ethnographic study of the Center for Comparative Effectiveness Research in Cancer Genomics (CANCERGEN). In this way, the analysis offers a novel empirical contribution at the nexus of these three domains (e.g. Birch 2016; Dussauge et al. 2015; Helgesson et al. 2016) and further situates the complexities of valuation practices in CER as an instructive site for tracing the broader implications they portend for the sociotechnical coordination of evaluation in healthcare more broadly. But as is evident from the interventions offered up in the foregoing chapters, this pushing of empirical boundaries is but a vehicle for a more significant project: refining the core conceptual apparatuses through which sociologists and students of science apprehend and appreciate, in all their complexity, the situations and artefacts with which the field confronts us.

2. Research summary

Each of the four preceding chapters was motivated by a distinct research question or *problématique*. In Chapter 2, we sought to trace the roots of CER and its historical and methodological synergies with the field of health technology assessment (HTA), highlighting four core themes that typified current comparative effectiveness research: a reappraisal of research designs; the reconfiguration of participation practices; issues of patient-centeredness and personalization; and questions of value assessment in healthcare. But in seeking to contribute to CER scholarship -- beyond the normative debates and prescriptive arguments about what CER *ought* to be, as were frequently found in the biomedical research and health policy literatures at the time -- we instead asked the question: What does CER look like in *practice*? In answering this question, we concluded by presenting the case of

the Center for Comparative Effectiveness Research in Cancer Genomics (CANCERGEN), an organization whose work was informed by the four aforementioned themes and thus surfaced as an exemplary instantiation of contemporary CER.

The scope of CANCERGEN's work consequently led to a consideration of the evaluation of personalized medicine technologies in oncology. As we argued, personalized medicine surfaced in the US amidst a nebulous regulatory environment and an onerous clinical trials system (e.g. Bourret et al. 2011; Hogarth 2015; Scoggins and Ramsey 2010) which posed challenges for how actors collectively went about performing (e)valuations of diagnostic tests in oncology. Two primary issues thus motivated the existence of CANCERGEN. On the one hand, many technologies were reaching the market without being required to demonstrate evidence of key clinical properties, including that their use results in improved patient outcomes. Meanwhile, efforts to leverage the publicly funded cancer clinical research infrastructure in redressing this evidence gap were hindered by federal funding cuts, increasing trial durations, and the premature closure of studies failing to accrue sufficient patient numbers needed to draw statistically meaningful conclusions. In response to the significant opportunity costs resulting from this crisis of allocative efficiency, CANCERGEN positioned itself as a living laboratory (cf. Bogner 2012) for testing and extending de novo methods for more effectively selecting and prioritizing diagnostic tests and study designs for conducting clinical trials.

Confronted with a unique *dispositif* that comprised the Center's approach to prioritizing and designing comparative effectiveness studies of personalized molecular diagnostic technologies in cancer care and prevention, we sought to examine the myriad (e)valuations that served as input and output of this process. Chapter 3 focused on the stakeholder engagement exercises deployed within this setting and asked how such a participatory format differs from other means of engaging in and with the world. During CANCERGEN's earlier phases, project organizers were mostly concerned with deliberative processes of stakeholder engagement, producing and publishing engagement frameworks and likening these exercises to examples of citizen engagement in the UK healthcare setting and to community-based participatory research in the US. But they also were quick to point out a key difference as well: stakeholders were said to have a vested interest in the design and outcomes of cancer clinical trials in a way that 'average citizens' do not. According to this logic, a

group of 13 such 'stakeholders' were recruited for the project – understood as representatives of various constituencies, including patients, consumers, clinicians, policymakers, and payers.

The group was first tasked with collectively deliberating and prioritizing an initial list of six precision diagnostic tests; through several rounds of voting and ranking, they were then to arrive at a penultimate top-three ranking; and, in the final phase, they were asked to come to a consensus on a single technology from the list with which to proceed in designing and conducting a large scale randomized controlled trial. The overall expectation was that trials designed in this new collaborative way – reflecting a vast array of viewpoints and interests – would result in the production of more robust evidence on the clinical usefulness of diagnostic technologies that could inform a diversity of decision-making spaces within the fragmented US healthcare system: from clinic, to industry, to the level of national policy. Approaching this process from the parallel perspectives of American pragmatist thought as well as French pragmatic sociology, we argued that *stakeholder engagement* implies a particular mode of coordination -- what Laurent Thévenot calls *engaging in plans* (e.g. 2007) which, that in the case of CANCERGEN, worked to limit the likelihood of an escalation to higher levels of critique -- what might be qualified a regime of *public justification* (e.g. Boltanski & Thévenot 2006).

Once the stakeholders had arrived at the shortlist of their three highest-ranked tests, emphasis within the project shifted from deliberation to *quantification* – and especially to the use of value of information analysis (VOI), a hybrid Bayesian method of economic-decision analytic modeling. Based on prior data gleaned from previously conducted clinical studies, as well as on other information including expert opinion, the models use standardized metrics such as the QALY (quality-adjusted life year) to quantitatively estimate the probable health improvements that will be brought about by more informative future evidence guiding the use of a given healthcare intervention when compared to current decision-making under existing uncertainty about that intervention. At a more general level, VOI is increasingly viewed as a promising avenue to more parsimonious decision-making in healthcare research – translating options for competing courses of action into pecuniary outputs, at once enabling (economic) valuation and commensuration (e.g. Espeland & Stevens 1998).

In one particular case, a team of CANCERGEN project organizers with expertise in VOI modeling produced a set of such models – one for each of the three remaining tests – and presented their results to the stakeholders. This, project organizers argued, was a way of making the uncertainties about the three technologies simply more 'explicit.' It did so by assigning a discrete dollar value to each test as an estimation of the value of uncertainty that conducting further research would ameliorate in each case. Organizers explained that if this 'value of research' exceeded the estimated cost of conducting the clinical trial, then the technology was deemed a legitimate candidate for a trial. But as all three tests fell into this latter category, the model results thus became a tool for stakeholders to conduct comparisons between the three remaining technologies, presented in the form of bar graph depicting the 'value' of each test side by side, with the monetary sum listed along the graph's x-axis as if to say: 'Here's where you get the best bang for your research buck!'

Taking these modeling practices as its object of inquiry, Chapter 4 probed into the nature of economic numbers which are produced within a regime of planned engagement (outlined previously in Chapter 3). Here we argued that while the outputs of VOI models were rendered as pecuniary values within the CANCERGEN setting, these numbers were not coextensive with *market prices*, nor were they qualified according to any other publicly legible order of worth (Boltanski & Thévenot 2006). Instead, drawing from Lampland's (2010) work on formalizing practices and Thévenot's (2015) notion of certification, we showed that VOI functions as a mode of *provisional economization* which serves to *certify the value* of options to be chosen within a circumscribed planning exercise.

Chapter 5 drew together many of the threads of the previous three chapters in order to understand how this convergence of qualitative and quantitative valuations proved propitious to answering the question: 'Which technology to trial?' -- that is, the processes which led to the selection of a single diagnostic technology with which to proceed in conducting a large scale CER trial. In particular, we traced how over the course of multiple rounds of stakeholder deliberations and voting, a set of breast cancer tumor markers went from being barely visible as a priority for the stakeholder group to eventually becoming the central focus of CANCERGEN's work in designing a study to test the clinical utility of these technologies. Returning again to Thévenot's notion of *engaging in plans*, this chapter demonstrated how even within an ostensibly single 'regime' there nevertheless exists the possibility for a plurality of modes of valuation -- and yet they resisted being deployed

simultaneously, rather unfolding over a series of *evaluative moments* which punctuated CANCERGEN's three year existence.

2. Originality of contributions & sociological implications

This work has made several original contributions (Guetzkow et al. 2004) to scholarship in medical sociology, STS, and valuation studies -- resting along both empirical and theoretical/conceptual axes -- with corresponding implications for how these fields apprehend and theorize (e)valuation, participation, and quantification.

With regard to the empirical axis, as argued above, this thesis is the first and only ethnographic study of the emergence and practices of comparative effectiveness research. Moreover, in focusing on the case of CANCERGEN, an additional contribution here is that the preceding analysis has moved the analysis of knowledge production and decision-making in biomedical research planning much further upstream than what has been described heretofore. Sociology and STS have done well to describe the judgments and (e)valuations that go into designing, recruiting for, analyzing, and synthesizing clinical trials and clinical trial data (Epstein 2008; Helgesson & Lee 2017; Moreira 2005). However, up until now scant ethnographic attention has been paid to more foundational questions regarding which technologies to subject to clinical study in the first place, and to the practices which seek to establish the legitimacy of these decisions.

In the latter case, I have demonstrated a novel conceptual and theoretical approach to the sociological study of (e)valuation -- and to participatory engagements and quantification practices in particular -- which affords a more faithful accounting of these phenomena compared to many of the existing approaches in the field. For instance, in pointing to a generalized tendency within contemporary scholarship to frame participation as being of a necessarily 'public' kind, I have deployed the case of CANCERGEN as something of a counter-example, showing how it instead constitutes what I have called a *pre-public platform*. Here, I call attention to the fact that the participation processes I observed in my fieldwork were not oriented by a particular vision of the common good, nor did they aim to sustain an open-ended process for collectively seeking solutions to complex societal problems -- both of which are hallmarks of such 'public' situations (cf. Boltanski & Thévenot 2006; Callon et al. 2009; Dewey 1946). Rather, I point to a certain mode of participation

oriented towards a much more limited good of achieving a stated objective. what Thévenot (2007) refers to as *engaging in planned action*.

In this planning regime, the legitimacy of participants is evaluated not on the basis of their being publicly-qualified 'citizens' but rather as *stakeholders* who are capable of expressing their 'interests' and 'opinions' in a functionally-equipped environment that is sequestered from the public sphere. That such an engagement may nevertheless escalate into a properly public dispute at some future moment -- for example by having actors denounce the format for being exclusionary and waging critiques according to higher principles of justice -- implies that it remains *prior* to the public so long as actors make no attempt to translate their concerns to more generalizable claims for justice or move to question the underlying conditions of the exercise. Moreover, that these participation exercises might be funded by public grants; that they seek to make decisions that impact the further distribution of public funds; and that those funds might have an impact on public health – none of this implies that the engagement itself is of a public kind – a confusion one finds all too frequently made in the STS literature. My explorations add considerable empirical and conceptual depth to ongoing sociological debates about the nature of participation, and help further disentangle some of the key conceptual difficulties that I have argued often lead to slightly misguided or misplaced conclusions.

My analysis also adds texture to existing scholarship in economic sociology and the anthropology of markets, particularly those focused on themes of economic valuation. As I have described above, one use of value of information analysis (VOI) in CANCERGEN was to produce a discrete economic value for each of the three shortlisted candidate technologies that were being considered for subjecting to the rigors of a large-scale comparative effectiveness trial. In unpacking how these models were developed, deployed, and presented to the External Stakeholder Advisory Group I highlighted that the pecuniary outputs of VOI modeling in CANCERGEN were based on certain calculative conventions. For instance, the health gains calculated in these VOI models were transformed into a dollar value by applying a willingness to pay (WTP) threshold; this conventional conversion rate states what the 'average' citizen of a given country is willing to pay for a year of life in a state of perfect health, but one whose methodological underpinnings have been the subject of some doubt within the field of health economics (cf. Grosse 2008; Neumann et al. 2014; Ryen &

Svensson 2014). Accordingly, rather than accepting that WTP reflected any 'truth to nature' (cf. Centemeri 2012; Daston & Galison 2007), CANCERGEN project organizers approached WTP as a 'shorthand' or 'rule of thumb,' relying on the capacities of these numbers for facilitating commensuration and visualizing differential value across multiple competing options for trial investment.

Comparing this to a subsequent deployment of VOI outwith the CANCERGEN context, I then moved on to claim that economized VOI outputs developed within CANCERGEN were neither *prices* nor any other *market* number -- derived, for instance, from market behavior and subsequently applied to the cost or value of a commodity; nor were they any other *market*-qualified number (e.g. Caliskan 2007; Fourcade 2011; Maldonado 2016; Muniesa 2007; Pallesen 2015). Indeed, VOI outputs appeared to not correspond with any publicly legible order of worth (Boltanski & Thévenot 2006). Rather, in developing the aforementioned notions of *provisional economization* and *certified worth*, I have shown that the primary reference for VOI analysis in CANCERGEN was the more proximal setting of planning and prioritizing clinical studies. Thus, in a way similar to how I framed stakeholder engagement as occurring in a moment prior to public situations, so too did the economized outputs of VOI models exist as a type of pre-public economic number. That is, while they might eventually come to impact economic decision-making in the public sphere at some later moment, they did not serve this function in the more immediate instance under investigation.

Finally, applying this insight symmetrically to both participation and quantification practices that were rolled out under the CANCERGEN banner, I conclude that sociological analysis is well suited to consider all evaluative activity as unfolding within an *ecology of engagements and (e)valuations* that enables the analyst to trace coordination practices and sociotechnical artefacts in their waxing and waning from moments of publicity.

3. Limitations

One limitation of this study has to do with methodology and research design. While I have relied on a combination of participant-observation, in-depth interviews, and analyses of the the myriad of publications and internal documents produced by CANCERGEN project participants. I was unable to be everywhere all the time, and there were numerous meetings and phone calls that I was unaware of or unable to participate in over the course of my fieldwork with the project. In this way, my

analysis is possibly constrained by the faithfulness with which the available resources -- especially meeting transcripts and summary documents circulated among project participants -- represent what actually happened in these instances that I was not able to observe firsthand.

Furthermore, as an ethnographic study primarily focused on the singular case of CANCERGEN, a second limitation of this thesis is that it has not accounted for the many other possible ways of coordinating comparative effectiveness research. As I have explained in the introductory chapter, my research was in fact initially envisaged as a multi-sited ethnography that specifically sought to examine CER in all of its multiplicity across multiple local sites where it was enacted. Hampered primarily by practical contingencies, however, such comparative work was left by the wayside and so many questions remain unanswered as to whether the modes of (e)valuation carried out in these other sites correspond with or otherwise trouble the conclusions I have reached with regard to CANCERGEN's own work and styles of organizing.

A related limitation is that, even were the aforementioned comparative study to have been carried out, it would still be limited to a series of *local* instantiations, leaving other scales of comparative effectiveness research unexamined. For instance, as discussed in Chapter 2, CER was formalized at the national level through the formation of the Patient Centered Outcomes Research Institute which also placed stakeholder engagement front and center in its activities. Whether national institutes like PCORI's own modes of stakeholder engagement demonstrate similar qualities of confinement that we found in CANCERGEN, or if they are somehow more propitious to public debate and critique, remains an unanswered question as well.

A third and final possible limitation of this research is the restricted time horizon which it covers. We have seen that CANCERGEN was funded for three years, while I have also suggested (e.g. in Chapter 4) that elements of work begun during this time continued in other settings following the sunsetting of its funding period. Thus, while we were able to follow many of the evaluative activities over the course of the three years during which I conducted my fieldwork, a more comprehensive longitudinal perspective on how the various evaluative apparatuses laid out in the previous chapters evolved -- and indeed continue to evolve -- is missing from the present analysis. This limits our ability to test the durability of my conclusions relating to the aforementioned notion of an *ecology of engagements and (e)valuations* which posits a temporal flux of participatory and quantification practices.

Unlike CANCERGEN, PCORI continues to operate as the national organizing body for CER in the United States and thus so too is there a missing narrative about the enduring features of comparative effectiveness research as enshrined in national policy and an account of the myriad changes that have taken place between the organization's inception in 2009 and what CER looks like in today's perhaps more turbulent political climate.

4. Future directions for research

There exist a number of possible avenues for further pursuing the empirical and theoretical contributions advanced in the preceding chapters, only a few of which I will touch upon in this concluding section.

Value of information analysis was not something I had set out to study at the outset of this project and yet by the end it had come to occupy a great deal of my intellectual attention. Because of the empirical limitations discussed above, the scope of this thesis only allowed me to attend to its more immediate deployment within the CANCERGEN context. Yet as I briefly noted in Chapter 4, much of methodological refinement for using VOI in healthcare and biomedicine has been carried out in the United Kingdom, whose National Health Service administers a single-payer system of healthcare provision. The subsequent adaptation of the method for prioritizing technologies and clinical trials in the United States meant that there were many elements of the approach which needed to be translated (Latour 1987) into this new context. Looking more closely at how VOI has been developed and deployed in the UK would doubtless shed additional light on the precise nature of these more recent translations. This, in turn, affords comparative leverage in grasping how evaluation is carried out in different jurisdictions (here, the UK and the US), each of which sees distinct convergences of health economists and their tools, national policy infrastructures, and cultural repertoires of healthcare delivery (e.g. Ashmore et al. 1989; Hirschman & Berman 2014).

Another opportunity for exploring the ongoing translations of (e)valuative technologies would be to return to the national clinical trials infrastructure, and SWOG in particular, which continues to be a site of experimentation for VOI analysis. While Chapter 4 suggests that the method was subsequently adapted for a new group of stakeholders following CANCERGEN -- in this case, members of the various SWOG disease committees who arbitrate whether to send study proposals up the decision-making chain (Bennette et al. 2016) -- it appears that still further work is being

carried to determine the tractability of VOI among SWOG leadership itself. Interestingly, the authors of a paper detailing this most recent 'experiment' note: 'Although there was general acceptance of the VOR methodology and appreciation of its potential value for decision making, the study investigator team did find ongoing resistance to VOR from a minority of investigators' who argued that 'the clinical impact of the drugs [being trialed] superseded the economic impact' and so the VOI team responded by creating a 'clinical VOR' (Carlson et al. 2018:8). But even after modifying this approach such that 'SWOG's decision-makers and their VOR preference heterogeneity' were taken into account, the authors note persistent resistance to VOI among a group of SWOG stakeholders, who claimed that the method in fact *hindered* the evaluation process and several of whom rejected the idea of adding VOI as an element of SWOG's study proposal evaluation process altogether (ibid.).

Further ethnographic work in this area promises to improve our knowledge about how particular types of information is formatted in (non)economic terms, according to certain logics, and how these multiple and oftentimes competing registers of value are negotiated and (un)settled in determining how precious public resources are allotted. This is especially important given the US Institute of Medicine's[10] calls for developing processes that improve the return on 'research investments' in cancer clinical trials (ibid.; Nass et al. 2010).

Following Muniesa and colleagues (2017), this may signal something beyond research studies simply 'becoming just an object of funding,' primarily requiring a demonstration of 'needs and worthwhileness' (129). Rather, we may instead be seeing a process of turning trials into *assets* via a process of *assetization* which 'establish[es] a rapprochement between disparate things (a bundle of rights, a living body, a piece of waste, a sum of money, a graphic document, a human disease, a microscopic particle, you name it) under the banner of this becoming investment' (ibid.:130). As compared with *commodities*, whose value the STS literature has frequently taken as intrinsic or latent, the value of assets 'becomes the value of investing in that thing' and enacted as 'prospective return' -- that is, the subject of valuation processes rather than a reified *object* of value (Muniesa et al. 2017:129 cf. Birch 2016:462). In this way, while I have characterized (and tacitly defended) CANCERGEN's VOI exercises as a mode of provisional economization -- and thus perhaps

[10] As of July 2015, The IOM has since been renamed the National Academy of Medicine

immune from criticisms that it constitutes a method of putting a 'market price' on what should remain a precious public good (cf. Fourcade 2011) -- zooming out from this more immediate setting reveals that such experiments may well be indicative of a much more far-ranging process of 'encoding' these public goods as assets, which threatens to to privilege the intentions of financiers (even if public institutions) at the expense of other voices and concerns (Muniesa et al. 2017:130).

Whether this is in fact what is occurring amidst the proliferation of VOI into ever higher levels of evaluation and decision-making in US cancer clinical research remains an unanswered empirical question. If nothing else, though, it provides fertile ground for exploring new empirical terrain and for redeploying and further testing the durability of the conceptual and theoretical innovations I have presented in this thesis regarding the flux of publicity and the scale and scope of evaluations that are carried out along the way.

5. References

Ashmore, M., Mulkay, M. J., & Pinch, T. J. (1989). *Health and efficiency: A sociology of health economics*. Open University.

Bennette, C. S., Veenstra, D. L., Basu, A., Baker, L. H., Ramsey, S. D., & Carlson, J. J. (2016). Development and Evaluation of an Approach to Using Value of Information Analyses for Real-Time Prioritization Decisions Within SWOG, a Large Cancer Clinical Trials Cooperative Group. *Medical Decision Making*, *36*(5), 641–651.

Birch, K. (2017). Rethinking Value in the Bio-economy: Finance, Assetization, and the Management of Value. *Science, Technology, & Human Values*, *42*(3), 460–490.

Bogner, A. (2012). The paradox of participation experiments. *Science, Technology & Human Values*, *37*(5), 506–527.

Boltanski, L., & Thévenot, L. (2006). *On Justification: Economies of Worth*. Princeton: Princeton University Press.

Bourret, P., Keating, P., & Cambrosio, A. (2011). Regulating diagnosis in post-genomic medicine: Re-aligning clinical judgment? *Social Science & Medicine*, *73*(6), 816–824.

Çalışkan, K. (2007). Price as a market device: cotton trading in Izmir Mercantile Exchange. *The Sociological Review*, *55*(s2), 241–260.

Callon, M., Lascoumes, P., & Barthe, Y. (2009). *Acting in an uncertain world: An essay on technical democracy*. Cambridge, MA: MIT Press.

Carlson, J. J., Kim, D. D., Guzauskas, G. F., Bennette, C. S., Veenstra, D. L., Basu, A., … Ramsey, S. D. (2018). Integrating value of research into NCI Clinical Trials Cooperative Group research review and prioritization: A pilot study. *Cancer Medicine*, *7*(9), 4251–4260.

Dewey, J. (1946). *The Public And Its Problems*. Chicago: Gateway Books.

Dussauge, I., Helgesson, C.-F., & Lee, F. (2015). *Value Practices in the Life Sciences and Medicine*. Oxford: Oxford University Press.

Epstein, S. (2008). The Rise of Recruitmentology: "Clinical Research, Racial Knowledge, and the Politics of Inclusion and Difference." *Social Studies of Science*, *38*(5), 801–832.

Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a Social Process. *Annual Review of Sociology*, *24*(1), 313–343.

Fourcade, M. (2011b). Price and Prejudice: On Economics, and the Enchantment/Disenchantment of Nature. In J. Beckert & P. Aspers (Eds.), *The Worth of Goods. Valuation and Pricing in the Economy*. Oxford, England: Oxford University Press.

Grosse, S. D. (2008). Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold. *Expert Review of Pharmacoeconomics & Outcomes Research*, *8*(2), 165–178.

Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is Originality in the Humanities and the Social Sciences? *American Sociological Review*, *69*(2), 190–212.

Helgesson, C.-F., & Lee, F. (2017). Valuations as Mediators Between Science and the Market: How Economic Assumptions Shape Pharmaceutical Trial Designs. *Science as Culture*, *26*(4), 529–554.

Helgesson, C.-F., Lee, F., & Lindén, L. (2016). Valuations of experimental designs in proteomic biomarker experiments and traditional randomised controlled trials. *Journal of Cultural Economy*, *9*(2), 157–172.

Hirschman, D., & Berman, E. P. (2014). Do economists make policies? On the political effects of economics. *Socio-Economic Review*, mwu017.

Hogarth, S. (2015). Neoliberal technocracy: Explaining how and why the US Food and Drug Administration has championed pharmacogenomics. *Social Science & Medicine*, *131*, 255–262.

Lampland, M. (2010). False numbers as formalizing practices. *Social Studies of Science*, *40*(3), 377–404.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Maldonado, O. J. C. (2017). Price-effectiveness: pharmacoeconomics, value and the right price for HPV vaccines. *Journal of Cultural Economy*, *10*(2), 163–177.

Moreira, T. (2005). Diversity in clinical guidelines: the role of repertoires of evaluation. *Social Science & Medicine*, *60*(9), 1975–1985.

Muniesa, F. (2007). Market technologies and the pragmatics of prices. *Economy and Society*, *36*(3), 377–395.

Muniesa, F., Doganova, L., Ortiz, H., Pina-Stranger, Á., Paterson, F., Bourgoin, A., … Yon, G. (2017). *Capitalization: A Cultural Guide*. Paris: Presses des Mines.

Nass, S. J., Moses, H. L., & Mendelsohn, J. (Eds.). (2010). *A national cancer clinical trials system for the 21st century: reinvigorating the NCI Cooperative Group Program*. Washington, DC: National Academies Press.

Neumann, P. J., Cohen, J. T., & Weinstein, M. C. (2014). Updating cost-effectiveness—the curious resilience of the $50,000-per-QALY threshold. *New England Journal of Medicine*, *371*(9), 796–797.

Pallesen, T. (2015). Valuable Assemblages - Or Assembling Values. In M. Kornberger, L. Justesen, J. Mouritsen, & A. K. Madsen (Eds.), *Making Things Valuable* (pp. 126–147). Oxford: Oxford University Press.

Ryen, L., & Svensson, M. (2015). The Willingness to Pay for a Quality Adjusted Life Year: A Review of the Empirical Literature. *Health Economics*, *24*(10), 1289–1301.

Scoggins, J. F., & Ramsey, S. D. (2010). A national cancer clinical trials system for the 21st century: reinvigorating the NCI Cooperative Group Program. *Journal of the National Cancer Institute*, *102*(17), 1371–1371.

Thévenot, L. (2007). The plurality of cognitive formats and engagements moving between the familiar and the public. *European Journal of Social Theory*, *10*(3), 409–423.

------. (2015). Certifying the World: Power Infrastructures and Practices in Economies of Conventional Forms. In P. Aspers & N. Dodd (Eds.), *Re-Imagining Economic Sociology* (pp. 195–216). Oxford: Oxford University Press.

Appendix A: Research Ethics Board Approvals

McGill

Research Ethics Board Office
McGill University
1555 Peel Street, 11th floor
Montreal, QC H3A 3L8

Tel: (514) 398-6831
Fax: (514) 398-4644
Ethics website: www.mcgill.ca/researchoffice/compliance/human/

Research Ethics Board I
Certificate of Ethical Acceptability of Research Involving Humans

REB File #: 304-0209

Project Title: Gold standards, silver linings: socio-technical elements of cancer treatment

Principal Investigator: Andrew S. Hoffman **Department:** Sociology

Status: Ph.D. student **Supervisor:** Prof. Alberto Cambrosio

This project was reviewed on June 1, 2009 by

Expedited Review ✓
Full Review ____

Elaine Weiner, Ph.D.
Vice-Chair, REB I

Approval Period: June 1, 2009 **to** May 31, 2010

This project was reviewed and approved in accordance with the requirements of the McGill University Policy on the Ethical Conduct of Research Involving Human Subjects and with the Tri-Council Policy Statement: Ethical Conduct For Research Involving Humans

---

*All research involving human subjects requires review on an annual basis. A Request for Renewal form should be submitted 2-3 weeks before the above expiry date.
*If a project has been completed or terminated and ethics approval is no longer required, a Final Report form must be submitted.
*Should any modification or other unanticipated development occur before the next required review, the REB must be informed and any modification can't be initiated until approval is received.

McGill

Research Ethics Board Office
McGill University
1555 Peel Street, 11th floor
Montreal, QC H3A 3L8

Tel: (514) 398-6831
Fax: (514) 398-4644
Ethics website: www.mcgill.ca/researchoffice/compliance/human/

Research Ethics Board I
Certificate of Ethical Acceptability of Research Involving Humans

REB File #: 16-0610

Project Title: Personal-izing Standards: Personalized Medicine, Clinical Practice Guidelines, and the Advent oif Comparative Effectiveness Research

Principal Investigator: Andrew S. Hoffman

Department: Sociology

Student Status: Ph. D. Student

Supervisor: Prof. A. Cambrosio

Funding Agency/Title: SSHRC

This project was reviewed on 3 June 2010 by

Expedited Review ___x___
Full Review _____

George Wenzel

George Wenzel, Ph.D.
Chair, REB I

Approval Period: June 03 2010 to June 02, 2011

This project was reviewed and approved in accordance with the requirements of the McGill University Policy on the Ethical Conduct of Research Involving Human Subjects and with the Tri-Council Policy Statement: Ethical Conduct For Research Involving Humans.

---

* All research involving human subjects requires review on an annual basis. A Request for Renewal form should be submitted 2-3 weeks before the above expiry date.
* When a project has been completed or terminated a Final Report form must be submitted.
* Should any modification or other unanticipated development occur before the next required review, the REB must be informed and any modification can't be initiated until approval is received.

McGill University

ETHICS REVIEW
RENEWAL REQUEST/STUDY CLOSURE FORM

Continuing review of research involving humans requires, at a minimum, the submission of an annual status report to the REB. This form must be completed to request renewal of ethics approval. If a renewal is not received before the expiry date, the project is considered no longer approved and no further research activity may be conducted. When a project has been completed, this form can also be used to officially close the study. To avoid expired approvals and, in the case of funded projects, the freezing of funds, this form should be returned 2-3 weeks before the current approval expires.

REB File #: 16-0610
Project Title: 'Personalizing Standards: The Advent of Comparative Effectiveness Research in the Postgenomic Era'
Principal Investigator: Andrew S. Hoffman
Department/Phone/Email: Sociology/Social Studies of Medicine; (347) 881 6340; andrews.hoffman@mail.mcgill.ca
Faculty Supervisor (for student PI): Dr. Alberto Cambrosio

1. Were there any significant changes made to this research project that have any ethical implications? ___Yes _X_No
   If yes, describe these changes and append any relevant documents that have been revised.

2. Are there any ethical concerns that arose during the course of this research? ___ Yes _X_ No. If yes, please describe.

3. Have any participants experienced any adverse events in connection with this research project? ___ Yes _X_ No
   If yes, please describe.

4. Is this a funded study? __X_ Yes ____No. If yes, list the agency name and project title and the Principal Investigator of the award if not yourself. This information is necessary to ensure compliance with agency requirements and to avoid any interruption in funds.

Funding for the study 'Personal-izing Standards: Personalized Medicine, Clinical Practice Guidelines, and the Advent of Comparative Effectiveness Research'
--McGill University Internal SSHRC Award (2010-2011)
--CIHR STIHR in Health Care, Technology, and Place Doctoral Fellowship (2010-2011)
--APOGEE/CanGeneTest Knowledge Network in Genomics Doctoral Fellowship (2010-2012)

__X___ Check here if this is a request for renewal of ethics approval.

_____ Check here if the study is to be closed and continuing ethics approval is no longer required. A study can be closed when all data collection has been completed and there will be no further contact with participants.

Principal Investigator Signature: _____ Date: March 17, 2011

Faculty Supervisor Signature: _____ Date: March 15, 2011
(for student PI)

Submit to Lynda McNeil(lynda.mcneil@mcgill.ca), Research Ethics Officer, James Admin. Bldg.,rm429 845 Sherbrooke Street West, Montreal, QC H3A 2T5 tel:514-398-6831 fax: 514-398-4644. Electronic submissions with scanned signatures are accepted but must come from the PI's McGill email. (version 11-2010)

For Administrative Use **REB:** X REB-I ___ REB-II ___ REB-III

___ The closing report of this terminated project has been reviewed.

X The continuing review for this project has been reviewed and approved.

X Expedited Review ___ Full Review

Signature of REB Chair or designate: L McNeil Date: April 1, 2011

Approval Period: April 1, 2011 to March 31, 2012

Submit to Lynda McNeil(lynda.mcneil@mcgill.ca), Research Ethics Officer, James Admin. Bldg.,rm429 845 Sherbrooke Street West, Montreal, QC H3A 2T5 tel:514-398-6831 fax: 514-398-4644. Electronic submissions with scanned signatures are accepted but must come from the PI's McGill email.

(version 11-2010)

McGill University

ETHICS REVIEW
RENEWAL REQUEST/STUDY CLOSURE FORM

Continuing review of research involving humans requires, at a minimum, the submission of an annual status report to the REB. This form must be completed to request renewal of ethics approval. If a renewal is not received before the expiry date, the project is considered no longer approved and no further research activity may be conducted. When a project has been completed, this form can also be used to officially close the study. To avoid expired approvals and, in the case of funded projects, the freezing of funds, this form should be returned 2-3 weeks before the current approval expires.

REB File #: 16-0610
Project Title: 'Personalizing Standards: The Advent of Comparative Effectiveness Research in the Postgenomic Era'
Principal Investigator: Andrew S. Hoffman
Department/Phone/Email: Sociology/Social Studies of Medicine; (347) 881 6340; andrews.hoffman@mail.mcgill.ca
Faculty Supervisor (for student PI): Dr. Alberto Cambrosio

1. Were there any significant changes made to this research project that have any ethical implications? ___Yes _X_No
   If yes, describe these changes and append any relevant documents that have been revised.

2. Are there any ethical concerns that arose during the course of this research? ___ Yes _X_ No. If yes, please describe.

3. Have any participants experienced any adverse events in connection with this research project? ___ Yes _X_ No
   If yes, please describe.

4. Is this a funded study? _X_ Yes ____No. If yes, list the agency name and project title and the Principal Investigator of the award if not yourself. This information is necessary to ensure compliance with agency requirements and to avoid any interruption in funds.

Funding for the study 'Personal-izing Standards: Personalized Medicine, Clinical Practice Guidelines, and the Advent of Comparative Effectiveness Research'
--McGill University Internal SSHRC Award (2010-2011)
--CIHR STIHR in Health Care, Technology, and Place Doctoral Fellowship (2010-2011)
--APOGEE/CanGeneTest Knowledge Network in Genomics Doctoral Fellowship (2010-2012)
--McGill Arts Graduate Research Travel Award (2011-2012)

__X__ Check here if this is a request for renewal of ethics approval.

_____ Check here if the study is to be closed and continuing ethics approval is no longer required. A study can be closed when all data collection has been completed and there will be no further contact with participants.

Principal Investigator Signature: _____ **Date:** 8 March 2012

Faculty Supervisor Signature: _____ **Date:** 8 March 2012
(for student PI)

Submit to Lynda McNeil(lynda.mcneil@mcgill.ca), Research Ethics Officer, James Admin. Bldg.,rm429 845 Sherbrooke Street West, Montreal, QC H3A 2T5 tel:514-398-6831 fax: 514-398-4644. Electronic submissions with scanned signatures are accepted but must come from the PI's McGill email. (version 11-2010)

For Administrative Use **REB:** X REB-I ___ REB-II ___ REB-III

___ The closing report of this terminated project has been reviewed.

__X__ The continuing review for this project has been reviewed and approved.

__X__ Expedited Review ___ Full Review

Signature of REB Chair or designate: L McNeil Date: March 15, 2012

Approval Period: March 15, 2012 to March 14, 2013

Submit to Lynda McNeil(lynda.mcneil@mcgill.ca), Research Ethics Officer, James Admin. Bldg.,rm429 845 Sherbrooke Street West, Montreal, QC H3A 2T5 tel:514-398-6831 fax: 514-398-4644. Electronic submissions with scanned signatures are accepted but must come from the PI's McGill email.

(version 11-2010)

McGill University

ETHICS REVIEW
RENEWAL REQUEST/STUDY CLOSURE FORM

Continuing review of research involving humans requires, at a minimum, the submission of an annual status report to the REB. This form must be completed to request renewal of ethics approval. If a renewal is not received before the expiry date, the project is considered no longer approved and no further research activity may be conducted. When a project has been completed, this form can also be used to officially close the study. To avoid expired approvals and, in the case of funded projects, the freezing of funds, this form should be returned 2-3 weeks before the current approval expires.

REB File #: 16-0610
Project Title: 'Personalizing Standards: The Advent of Comparative Effectiveness Research in the Postgenomic Era'
Principal Investigator/Department: Andrew S. Hoffman
Email: andrews.hoffman@mail.mcgill.ca
Faculty Supervisor (if student is the PI): Alberto Cambrosio (alberto.cambrosio@mcgill.ca)

1. Were there any significant changes made to this research project that have any ethical implications? ___Yes __X_No
   If yes, and these have not already been reported to the REB, describe these changes and append any relevant documents that have been revised.

2. Are there any ethical concerns that arose during the course of this research? ___ Yes _X_ No. If yes, please describe.

3. Have any participants experienced any unanticipated issues or adverse events in connection with this research project?
   ___ Yes _X_ No
   If yes, please describe.

4. Is this a funded study? ___Yes _X_ No.
   List the agency name and project title and the Principal Investigator of the award if not yourself. This information is necessary to ensure compliance with agency requirements and that there is no interruption in funds.

_X_ Check here if this is a **request for renewal** of ethics approval.

____ Check here if the **study is to be closed** and continuing ethics approval is no longer required. A study can be closed when all data collection has been completed and there will be no further contact with participants.

Principal Investigator Signature: _____ Date: 13 March 2013

Faculty Supervisor Signature: _____ Date: March 13, 2013
(if PI is a student)

For Administrative Use **REB:** _X_REB-I ___REB-II ____REB-III

____ The closing report of this terminated project has been reviewed

_X_ The continuing review for this project has been reviewed and approved

_X_ Delegated Review ____Full Review

Signature of REB Chair or designate: _____ Date: March 15, 2013

Approval Period: March 15, 2013 to March 14, 2014

Submit to Lynda McNeil(lynda.mcneil@mcgill.ca), Research Ethics Officer, James Administration Building, 845 Sherbrooke Street West suite 429, Mtl., QC H3A0G4; fax: 398-4644 tel: 398-6831. Electronic submissions with scanned signatures are accepted but must come from the PI's McGill email.

(version 12/12)

McGill University
ETHICS REVIEW
RENEWAL REQUEST/STUDY CLOSURE FORM

Continuing review of research involving humans requires, at a minimum, the submission of an annual status report to the REB. This form must be completed to request renewal of ethics approval. If a renewal is not received before the expiry date, the project is not considered to be approved and no further research activity may be conducted. When a project has been completed, this form can also be used to officially close the study. To avoid expired approvals and, in the case of funded projects, the freezing of funds, this form should be returned 2-3 weeks before the current approval expires.

REB File #: 16-0610
Project Title: Personalizing standards: personalized medicine, clinical practice guidelines and the advent of comparative effectiveness research
Principal Investigator/Department: Andrew S. Hoffman, Social Studies of Medicine/Sociology
Email: andrews.hoffman@mail.,mcgill.ca
Faculty Supervisor (if student is the PI): Dr. Alberto Cambrosio

1. Were there any significant changes made to this research project that have any ethical implications? ___Yes _X_No
   If yes, and these have not already been reported to the REB, describe these changes and append any relevant documents that have been revised.

2. Are there any ethical concerns that arose during the course of this research? ___ Yes __X_ No
   If yes, please describe.

3. Have any participants experienced any unanticipated issues or adverse events in connection with this research project?
   ___ Yes _X_ No
   If yes, please describe.

4. Is this a funded study? ___Yes _X_No.
   If yes, list the agency name and project title and the Principal Investigator of the award if not yourself. This information is necessary to ensure compliance with agency requirements and that there is no interruption in funds.

5. Did this project require REB approval from another Institution/Board? ___Yes _X_No
   If yes, and the project is continuing, attach a copy of the current approval.

_X__ Check here if this is a **request for renewal** of ethics approval.

____ Check here if the **study is to be closed** and continuing ethics approval is no longer required. A study can be closed when all data collection has been completed and there will be no further contact with participants.

Principal Investigator Signature: _____ Date: 18 February 2014

Faculty Supervisor Signature: _____ Date: Feb. 19, 2014
(if PI is a student)

Submit to Lynda McNeil(lynda.mcneil@mcgill.ca), Research Ethics Officer, James Administration Building, 845 Sherbrooke Street West suite 429, Mtl., QC H3A0G4; fax: 398-4644 tel: 398-6831. Electronic submissions with scanned signatures are accepted but must come from the PI's McGill email.

(version July 2013)

For Administrative Use **REB:** X REB-I ___REB-II ____REB-III

____ The closing report of this terminated project has been reviewed.

__X__ The continuing review for this project has been reviewed and approved. X Delegated Review ____ Full Review

Signature of REB Chair or designate: _____ Date: March 1, 2014

Approval Renewal Period: March 1, 2014 to Feb. 28, 2015

Submit to Lynda McNeil(lynda.mcneil@mcgill.ca), Research Ethics Officer, James Administration Building, 845 Sherbrooke Street West suite 429, Mtl., QC H3A0G4; fax: 398-4644 tel: 398-6831. Electronic submissions with scanned signatures are accepted but must come from the PI's McGill email.

(version July 2013)

Appendix B: Sample Research Consent Forms

McGill

Research Consent Form
McGill University

Gold Standards, Silver Linings: Socio-Technical Elements of Cancer Treatment Guidelines Production in the United States

P.I. Andrew S. Hoffman, PhD Student
Department of Sociology, McGill University
Ph: (514) 678-0106
E: andrews.hoffman@mail.mcgill.ca

Supervisor: Dr. Alberto Cambrosio, PhD
Chair, Social Studies of Medicine; Department of Sociology
PhL (514) 398-4981
E: alberto.cambrosio@mcgill.ca

Purpose of Research
The current study seeks to provide a more thorough understanding of clinical practice guidelines in the United States, particularly as they relate to cancer treatment. I am interested in the processes through which cancer treatment guidelines are created, and I wish to analyze both the various actors that are implicated in this work—including processes of systematic review and meta-analysis—as well as how this knowledge is deployed when professionals negotiate and arrive at consensus on the contents of a given guideline. The importance of this research lays in its explanatory power: with an increased understanding of what actors are involved in processes of guidelines development, how negotiations between these actors are structured, and what knowledges are ultimately incorporated into these guidelines, perhaps new configurations will emerge that can be of greater benefit to both patients' quality of life and the decision-making capabilities of practitioners

What is involved in participating?
You will be asked a few questions relating to your participation in work directly or indirectly related to the development of cancer treatment guidelines. The interview is intended to last one hour, but you may opt for a shorter or longer duration. Interviews will be audio taped and subsequently transcribed by the interviewer. The interviewer will be the only one with access to the interview materials. The data will not be published or written up in any way that may unintentionally disclose your identity unless you consent to having this information included; other individuals who you name in your account will remain confidential. However, even if you do not consent to having your name included, I cannot guarantee the complete confidentiality of your participation, as the organization with which you work is of a particular size that might enable potential readers of my final report to deduce who you are. The information gleaned in this study will ultimately be included in my doctoral dissertation, parts of which may be published in the form of books and/or scholarly journal articles.

Your participation in this study is voluntary and you may choose not to participate or withdraw at any time. You may also refuse to answer any questions without having to justify yourself. Should you at any point opt out of inclusion in this study, your interview transcript will be removed from my files. You will not be paid for your participation.

I have read the above terms and understand that my signature indicates free and informed consent to be interviewed as part of this study.

Do you agree to have this interview recorded: □ Yes □ No
Do you agree to have your identity included in this interview: □ Yes □ No

Respondent's Signature _____
Researcher's Signature _____ Date _____

Research Consent Form
McGill University

Personalizing Standards:
The Advent of Comparative Effectiveness Research in the Postgenomic Era

P.I. Andrew S. Hoffman, Ph.D. Student
Social Studies of Medicine, McGill University
Ph: (347) 881-6340
E: andrews.hoffman@mail.mcgill.ca

Supervisor: Dr. Alberto Cambrosio, PhD
Chair, Social Studies of Medicine
Ph: (514) 398-4981
E: alberto.cambrosio@mcgill.ca

Purpose of Research
The current study seeks to provide a more thorough understanding of comparative effectiveness research (CER) as a new evaluative technology for determining the effectiveness of medical interventions. I am particularly interested in the processes through which organizations are working to develop CER methodology, how CER attempts to deal with the proliferation new genomic technologies (and personalized medicine more broadly defined), as well as how CER results are being deployed in routine clinical practice (e.g. through their incorporation into clinical practice guidelines). The overall purpose of this research is to provide stakeholders with a wider lens through which to view these important changes in medical practice. In terms of what they mean for patients and health care practitioners, a more thorough understanding of such novel research practices will perhaps contribute to the formation of new configurations that can be of greater benefit to both patients' quality of life and the decision-making capabilities of practitioners. As well, this research will also address related concerns about clinical infrastructure, drug development, and regulatory norms that will significantly impact the way medicine is practiced in North America and around the world.

What is involved in participating?
You will be asked a few questions relating to your participation in work directly or indirectly related to the development of CER methodology, clinical practice guidelines, and/or the field of personalized medicine. The interview is intended to last one hour, but you may opt for a shorter or longer duration. Interviews will be audio taped and subsequently transcribed by the interviewer. The interviewer will be the only one with access to the interview materials, which will be stored on a password-protected computer. The data will be written up in such a way as to ensure confidentiality of the respondent unless you choose to be identified. The information gleaned in this study will ultimately be included in my doctoral dissertation, parts of which may be published in the form of books and/or scholarly journal articles.

Your participation in this study is voluntary and you may choose not to participate or withdraw at any time. You may also refuse to answer any questions without having to justify yourself. Should you at any point opt out of inclusion in this study, your interview transcript will be removed from my files. You will not be paid for your participation.

I have read the above terms and understand that my signature indicates free and informed consent to be interviewed as part of this study.

Do you agree to have this interview recorded: □ Yes □ No
Do you agree to have your identity included in this interview: □ Yes □ No

Respondent's Signature _____
Researcher's Signature _____ Date _____