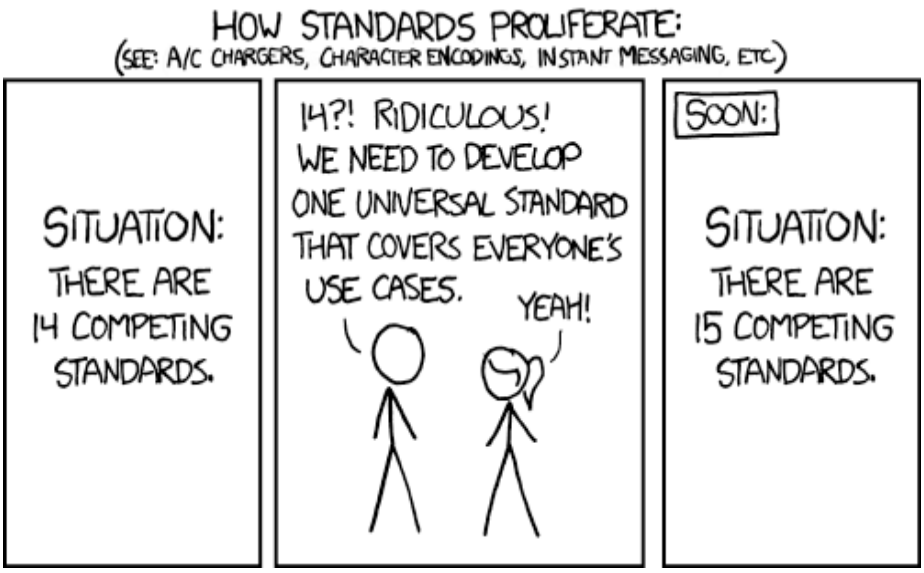
In search of standards that avoid standardization: The production and regulation of evidence based guidelines.

A. Loes Knaapen

Department of Sociology
McGill University
Montreal, Quebec, Canada

March 2013

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



By: XKCD Comics, www.xkcd.com/927/

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am very grateful to many people and organizations who have supported me during the research for, and writing of, this dissertation. First of all, the research would not have been possible without the curiosity, cooperation and time of the many guideline developers I observed, interviewed and interacted with. I would especially like to thank Beatrice Fervers, Jako Burgers and Melissa Brouwers for opening the doors of the SOR, CBO and PEBC guideline institutions and for allowing me to observe guideline development in action. Conversations with them and with Kitty Rosenbrand, Hans Messersmith and Sheila McNair, helped me understand the many aspects, projects, principles and people involved in, what, at first sight, seemed a bewildering world of guideline development. I am grateful to the Guidelines International Network working group on Patient and Public Involvement for welcoming my participant observation of their group, and I especially thank Antoine Boivin, Trudy van der Weijden, Carol Sakala and Corinna Schaeffer for inspiring and encouraging me. Co-chairing G-I-N PUBLIC was an energizing distraction during the writing of the dissertation.

I was privileged to write this dissertation under the supervision of Alberto Cambrosio. I am most grateful for his helpful insights, continuing mentorship, constructive commentary, and for providing encouragement and advice throughout. I am also thankful for the support and supervision I received from other faculty members during my time at the department of Social Science and Medicine. George Weisz and Thomas Schlich were essential in shaping my research interest and approach, helping me to imagine my dissertation project, and encouraging and challenging it while on its way. I have benefited from seminars, courses and mentorship by many others at McGill, I'd like to thank Axel van den Berg, Jennifer Fishman, Tobias Rees and Amélie Quesnel-Vallée.

I'm indebted to Patrick Castel and Hervé Cazeneuve for getting me started on the first part of the dissertation. I am grateful to Vololona Rabeharisoa and Madeleine Akrich, as well as Scott Greer and Paulette Kurzer, for inviting me to present my work and for giving me encouragement and commentary on earlier drafts.

Fellow graduate students in McGill University's Departments of Sociology, Anthropology and Social Studies of Medicine provided a great community that enlivened my academic and personal life. I am indebted to all my office mates (past and present) at SSOM for helping in so many small and large ways to bring this dissertation to a good end. In particular I thank Noémi Tousignant, Annalisa Salonius, Raul Necochea, Takeshi Uesugi, Pierre Minn (Lunch club rocks!), Emilio Dirlikov, Hanna Kienzler, Clarissa Allen and Hadi Karsoho for their friendship and support. Drew Hoffman and Dörte Bemme have helped me tremendously, reading, reviewing and editing various drafts and pieces of my writing. Special thanks must be given to Julieta Bach, who makes The Basement sparkle with her love and enthusiasm. Thanks are also due to administrative staff in the Sociology and SSOM departments, Daniela Arevalo, Heike Faerber, Adele Tarantino and Erin Henson, who helped me overcome administrative hurdles of graduate studies and research.

David Meltzer and Ellen Cohen at the University of Chicago welcomed me at the Center for Health and the Social Sciences (CHESS). Thank you so much, CHESS' office space provided a great environment for writing.

This research was made possible by financial support from the 61e Commission permanente de coopération franco-québécoise, from the Fonds de Recherche sur la Société et la Culture (Quebec), and a doctoral award from the Fonds de Recherche Santé Quebec. A CIHR Strategic Fellowship in Health Care Technology and Place (HCTP) at the University of Toronto provided me with financial support, and HCTP's mentors and fellows

provided intellectual support and encouragement to pursue this research. I am also indebted to the Canadian Institutes for Health Research (Institutes of Health and Social Policy Research) for their aid in funding to present this work at conferences.

Finally, it makes me so happy that my parents Sjef and Carla, and my sisters Simone and Marike, are proud of me, and support my adventures even when what I do is so ‘far from their beds’. I am grateful that my time at McGill has resulted in such wonderful friendships. Louigi & Rosie, Dörte and Karen, I am glad we will be (part-time) neighbors again. Finally, Andrew McCann is the best. I’m so lucky that the project we started in September 2006 will not be concluded with this dissertation, but is an experiment for life.

Abstract

This dissertation examines a novel type of standardization in medicine by investigating the production and regulation of clinical practice guidelines. As an important tool of the Evidence Based Medicine (EBM) movement, guidelines have been at the center of polarized debates in which they are praised for rationalizing medicine and criticized for undermining humanism in health care. Based on document analysis, interviews and (participant) observation, this dissertation examines how 'evidence based' guideline developers respond to such contradictory demands and critiques. In doing so, I present an empirical examination of the way EBM practices construct, avoid or reconcile tensions between supposed binaries such as universal/local, evidence/values, standards/individuality, science/care. After the introduction and a review of the existing literature on the epistemological basis and regulatory impact of EBM guidelines, the findings are presented in two parts. The first part analyzes how formal EBM principles are understood and performed during guideline production, with Chapter three highlighting the diversity of knowledge, procedures and materials required to classify Evidence and formulate the guideline text. Chapter four analyzes how an absence of Evidence is handled and proposes the term 'Evidence Searched Guidelines' to capture the distinctive character of EBM guidelines. The second part focuses on 'guidelines for guidelines' that regulate guideline development. Chapter five presents the Guidelines International Network, which aims for a 'universal' procedure for standard-setting, but objects to the standardization of guidelines. Chapter six presents four models of Patient & Public Involvement that aim to integrate evidence with 'values' and 'context', making guidelines more personalized, democratic, locally relevant and/or objective. In aiming for Gold Standards that avoid standardization, EBM guideline developers challenge the accounts of EBM proponents and critics alike. The contested relation between the standardized (or universal) and the individualized (or local) at the heart of Evidence Based Medicine is managed not by the quantity, quality or universality of evidence, but by standardization of procedures.

Résumé

Cette thèse étudie la production et la réglementation des lignes directrices pour la pratique clinique, afin d'examiner un type de standardisation médicale novateur. Servant d'outil important dans le mouvement de la médecine basée sur les données probantes (Evidence Based Medicine ou EBM), les lignes directrices ont été au centre de débats polarisés dans lesquels on fait l'éloge d'elles pour avoir rationalisées la médecine et on les critique d'avoir minées l'humanisme dans les soins de santé. Basée sur l'analyse de documents, des interviews et l'observation (participative), cette thèse examine comment les développeurs des lignes directrices répondent aux telles demandes et critiques contradictoires. Ce faisant, je présente un examen empirique des façons que les pratiques de la EBM construisent, évitent et réconcilient les tensions entre de présumés binaires, tels universel/local, données probantes/valeurs, standards/individualité et sciences/soins. Après l'introduction et l'analyse de la littérature existante sur les bases épistémologiques et l'impact de la réglementation des lignes directrices de la EBM, les résultats sont présentés en deux parties. La première partie analyse la compréhension et la performance des principes formels de la EBM durant la production des lignes directrices, le troisième chapitre mettant l'accent sur la diversité des connaissances, des procédures et des matériels nécessaires pour classifier les données probantes et formuler le texte des lignes directrices. Le quatrième chapitre analyse comment l'absence de données probantes est gérée et propose le terme «Evidence Searched Guidelines» (lignes directrices cherchantes les données probantes) pour saisir le caractère distinctif des lignes directrices de la EBM. La deuxième partie met l'accent sur les «lignes directrices pour les lignes directrices», qui régularisent le développement des lignes directrices. Le cinquième chapitre présente le Guidelines International Network, visant une procédure universelle pour l'établissement des normes, mais s'opposant à la standardisation des lignes directrices. Le sixième chapitre présente quatre modèles d'Implication du Public visant à intégrer les données probantes à des «valeurs» et des «contextes», afin de rendre les lignes directrices plus personnalisées, démocratiques, pertinentes et/ou objectives. Visant des Étalons-or qui évitent la standardisation, les développeurs des lignes directrices de la EBM mettent au défi à la fois les explications des défenseurs de la EBM et celles des détracteurs. Le rapport contesté entre standardisé (ou universel) et personnalisé (ou local) au cœur de la médecine fondée sur les données probantes n'est pas gouverné par la quantité, la qualité ou l'universalité des données probantes, mais par la standardisation des procédures.

Preface

The format of this dissertation is 'manuscript-based', as chapters three, four, five and six have been prepared as individual scholarly papers to be published as journal articles and book chapters. To allow each chapter to be comprehensible on its own, some overlap and repetition between the chapters exists. The collection of papers is the result of a unitary program of research, with a single overarching research question, design and data collection concerned with the production and regulation of guideline production according to the principles of Evidence Based Medicine. The introduction and literature review (in chapters one and two) as well as the concluding chapter clearly outline how each manuscript addresses a different aspect of the overall research question.

Contribution of Authors

For all chapters I am the primary author, and only chapter three has been co-authored. The authors' contributions are as followed Béatrice Fervers, Hervé Cazeneuve and Patrick Castel conceived of and designed the study, Loes Knaapen and HC gathered data, all authors advised on interpretation of data, LK and Alberto Cambrosio conducted primarily data analysis, LK wrote the first draft of the manuscript which PC and AC critically revised for important intellectual content, all authors reviewed and approved final manuscript. For all other chapters I am the sole author. The entire doctoral research was supervised by Alberto Cambrosio who advised on the research conception, design, analysis and interpretation of data, and provided important commentary on all manuscripts.

Statement of Original Scholarship

The work contained herein represents an original and important contribution to existing social science literature on the production of clinical practice

guidelines (CPGs), and the nature of and transformations in the evidence based medicine (EBM) 'paradigm'. To my knowledge, it is one of only two studies that use direct observation of guideline development in practice as its research method. By demonstrating the pragmatic and localized nature of guideline development and evidence assessment, this work challenges existing literature on the question 'what counts as evidence' in EBM (which are mostly normative and abstract critiques), as well as existing empirical studies of guideline development (that largely take the definition of evidence for granted, focusing instead on the regulatory impact of standards). Chapter three extends the existing analysis of guideline development 'in action' by providing specifications of how the production and justification of guideline text is achieved through the employment of institutional specific *dispositif* that coordinates experts, documents, statements, procedural rules and relations to the world external to the guideline development group. It has been published in Social Studies of Medicine (Knaapen et al., 2010). Chapter four is a unique empirical analysis of 'what counts as evidence', and what doesn't, in the actual construction of EBM tools. To better reflect the role of (non-)evidence in EBM guideline development, it coins the term 'evidence searched guidelines', and this chapter is under revision for publication in Social Studies of Science. Chapter five continues to investigate the procedural standards that are central to guideline production and regulation, now focusing on international collaboration. It is the first social science inquiry into the Guidelines International Network (GIN), a newly emerged international network that produces 'guidelines for guideline development' to regulate evidence based guidelines. It provides an original case study of international network building within the current era of Evidence Based Medicine marked by increased global knowledge and exchange. This chapter has been published in an edited volume on Europeanization of public health policy in the EU, as the network emerged as an attempt by the EU to establish a 'new governance' structure to

harmonize European clinical practice guidelines (Knaapen, 2012). Chapter six is more conceptual in nature, providing a comprehensive overview of existing rationales for Patient & Public Involvement (PPI) in guideline development. It draws on existing literature of models of PPI in other health care settings and initiatives, and enriches the classical political (democratic) and economic (consumer choice) rationales for PPI by highlighting a role for 'lay experts' in standard setting (drawing on Science & Technology Studies literature), and propose the novel role of 'Critical Witness' as a forth rationale for PPI in evidence based guideline development.

As a whole, the dissertation makes several larger claims. First, it argues that procedural standards are central to the 'EBMness' of guideline development, not the quantity, quality or universality of evidence. Second, the procedural standards that regulate guideline production legitimize diversity (by encouraging national variation in guidelines), accept pragmatic judgment and localized routines (by formalizing 'other considerations' and contextualized evidence assessment) and aim to include values (through patient involvement procedures). Third, by increasingly standardizing and globalizing the procedural rules, the much criticized guidelines (and guideline developers) gain legitimacy and professionalization, without globalizing the guidelines themselves.

CHAPTER 1 Introduction

The world of Evidence Based Guidelines

This dissertation investigates the production of a particular genre of treatment standards—called clinical practice guidelines —which are important instruments of Evidence Based Medicine (EBM). Clinical Practice Guidelines (CPGs), henceforth simply 'guidelines', are a particular type of medical standard that aims to regulate physicians' behavior through knowledge. Based on a synthesis of evidence from medical research (ideally randomized clinical trials) CPGs provide clinicians with recommendations for how best to treat a particular clinical condition. By providing a Gold Standard for clinicians to live up to, guidelines carry the promise of improving the quality of health care by simultaneously rationalizing and standardizing medical practices (Timmermans and Berg, 2003). As instruments of the Evidence Based Medicine paradigm, guidelines "become the tool of choice to weed out unwarranted variation in diagnostic or therapeutic practice and to enhance the scientific nature of medical care delivered" (Berg et al., 2000:766). Countless medical professional societies, public and governmental bodies, and private parties such as insurance companies have produced thousands of guidelines for clinical practice. Despite (or perhaps because) of such great promise and praise, both EBM and its guidelines have also been fiercely criticized. A rich literature in medicine, humanities and the social sciences has criticized EBM for how it defines evidence and how its guidelines reform medicine (Lambert, 2006; Mykhalovskiy & Weir, 2004). Summarizing the myriad critiques on Evidence Based Medicine, Mykhalovskiy and Weir conclude that "Discussions tend to be rather grand, organized as abstract critique of EBM rather than as empirical research of particular cases of its development or

use." (2004:1061). My dissertation responds to the need for more empirical studies on the actual workings of EBM and its instruments in daily practice, by investigating guideline development in practice, including the establishment of international meta-standards for guideline development.

Debating EBM guidelines: the nature of knowledge and regulatory power.

The existing literature on EBM guidelines falls generally within two kinds of debate. One is about the nature of 'knowledge' within the EBM movement, while the second focuses on the regulatory impact of guidelines on the medical profession.

The debate about knowledge is mostly conducted by philosophers, epidemiologists and medical researchers (Lambert, 2006). Critics object to the categorization of medical knowledge as a ranking of evidence, with 'basic research', opinion and case studies at the bottom and randomized clinical trials (RCTs) and systematic reviews of evidence at the top (see figure 1), calling out the "fixed hierarchy of evidence as the guilty source of [EBM's] questionable epistemic practices" (Goldenberg, 2009: 171). The automatic 'gold standard' status of RCTs is critiqued by pointing out the limitations and flaws of this trial methodology. The priority given to the narrow definition of

Figure 1. An Evidence Hierarchy.
Source: Evidence-Based Practice in the Health Sciences: Evidence-Based Nursing Tutorial. University of Illinois at Chicago, Information Services Department of the Library of the Health Sciences-Chicago.

evidence is criticized for its exclusionary nature, neglecting entire disciplines, research traditions and treatment domains and even the 'erasure of the patient' as unique individuals are seen to be replaced by standardized population statistics. The “politics of evidence” of the reductionist and exclusionary nature of the evidence hierarchy echoes the late Leigh Star’s claim that “one can read a surprising amount of social, political and philosophical context from a set of categories” (Bowker & Star, 2000:55). However, these discussions neglect that formal classifications do not determine practice, as “classificatory systems themselves can be understood to hide more than they reveal: they are reductions that efface the complexity and messiness of medical categorizing “ (Latimer et al., 2006:599). Instead of focusing exclusively on formal evidence categories and official EBM writings, this dissertation investigates the actual practices of producing and using evidence according to EBM. Novel insights into the nature of evidence can be gained by investigating what EBM guideline developers are actually doing, rather than what they ought to be doing (according to EBM's formal principles) or what they say they are doing (in interviews). Chapter three of this dissertation investigates the definition and classification of evidence as pragmatically achieved in practice, and emphasizes the institutionally specific rules, procedures and people necessary to formulate a guideline text. Chapter four investigates the incompleteness of evidence (both indeterminacy and absence) and the way ‘non-evidentiary’ justifications relate to evidence. I propose the term ‘evidence searched guidelines’ to reflect how EBM guideline developers follow the formal evidence classification and EBM rules, while maintaining the legitimacy of informal, local and pragmatic knowledge and justifications that fall outside that strict hierarchy.

The second debate in the extant literature focuses not on the epistemological basis of EBM guidelines, but rather concerns it self primarily with the regulatory impact of guidelines. The regulatory power of EBM

guidelines is assessed by drawing on classical sociological works on the medical profession's authority to control (perform, monitor, judge) medical work (Freidson, 1970; 1984), considering guidelines to shift the balance of countervailing professional, state, market and consumer powers within medicine (Light, 2000) or regulatory tools within the 'audit society' that emerged in the 1990s (Power, 1999; 2000). According to such frameworks guidelines are a tool used by external parties (typically private managers and public policy makers, but also patients, or specific 'elite' segments of the medical profession) to reduce professional power and skills of individual clinicians (Armstrong, 2007, 2002; McGivern and Fischer, 2012; Pope, 2003; Timmermans and Kolker, 2004). The detailed history of the emergence of guidelines shows that the narrative of guidelines as an external tool to control professional power and health care costs is too simplistic (Weisz et al., 2007). Some detailed empirical studies have shown the complexities of the regulatory impact of the development and use of guidelines, emphasizing that standards have to be embedded into the existing local (net)work, incorporating some of the skills, routines and resources they set out to replace (Timmermans and Berg, 2003). Furthermore by challenging existing intra-professional hierarchies and competition (Castel, 2009), or claiming a domain of professional expertise (Bowker and Star 2000; Berg et al., 2000), standards can both reduce and increase professional authority (Levay and Waks, 2009). Central to understanding the regulatory power of standards is the question: 'Who controls such a quality control tool?'. To address the issue of 'control over control' requires not only inquiry into the *use* of evaluation tools (Power, 2000), but also empirical analysis of who controls the *production* of such tools, as they determine *what* is made measurable, comparable and governable and what is left informal, flexible, ambiguous and invisible. To understand what is being standardized or evaluated in the first place, this dissertation analyses the *production* of guidelines, as well as how the quality control of guideline development is

organized. Chapter five presents the Guidelines International Network as a meta-regulatory network devoted to the quality control of guidelines, and chapter six investigates the role of patients and the public in the use and development of EBM guidelines.

A sociology of standardization

This dissertation adds novel insights to the nature of knowledge within the EBM movement and sheds light on novel types of (soft) regulation through standardization, moving the debate on EBM forward by bringing together epistemological and regulatory issues that are usually addressed separately. By considering guidelines as one element along a chain of textual *translations* linking knowledge production to the regulation of medical practices, I situate them as part of establishing a new kind 'regulatory objectivity' (Cambrosio, Keating, Schlich & Weisz, 2006). This dissertation thus provides ethnographic detail on regulatory objectivity in the specific domain of voluntary standards for clinical practice (distinct from laboratory, research or binding regulatory (i.e. FDA) standards (Cambrosio et al., 2009; Moreira, May & Bond, 2009; Hogle, 2009). In doing so, I aim to contribute to what Timmermans & Epstein (2010) have dubbed the emerging subfield of "sociology of standardization" that inquires into the political, financial, personal and epistemological issues at stake in standardization or lack thereof. What may at first seem a rather mundane topic, of key concern only to technocrats and engineers, sociologists of standardization argue should be of central concern to social scientists. By organizing and regulating our institutions, objects, behaviors and bodies, standards affect almost every aspect of our everyday lives: "Standards and classifications, however dry and formal on the surface are suffused with traces of political and social work." (Bowker & Star, 1996:214). Moreover, it proposes that social science inquiry goes beyond the typical, and contradictory, narratives associated

with standardization: "*Standards* are typically deemed laudatory; they are something one aspires to live up to. But *standardization* in its popular uses is derogatory; it connotes a dull sameness, the suppression of individuality in the service of industrial uniformity." (Timmermans & Epstein, 2010;71, my emphasis). Instead of treating standardization as inherently either good or bad, the authors call for careful empirical analysis of the myriad causes and consequences (sometimes unintended) of specific standards in distinct social domains. Empirical analysis of the *use* of standards in health care have shown that, inevitably, standards will not univocally bring quality (as promises claim) nor will standardization be fully 'dehumanizing' (as critics fear) (Timmermans & Berg, 2003). My doctoral research contributes to this literature by presenting empirical findings on the *production* and *regulation* of clinical practice guidelines within the domain of Evidence Based Medicine. Despite the rich normative literature that praises, criticizes and defends what EBM guideline production is and ought to be, almost no empirical accounts exist of how EBM guidelines are actually produced in current practice (Moreira, 2005 is a notable exception). By focusing on the practice of guideline development amidst the contradictory demands and critiques posed on standard setting, my work seeks to answer the question *How do guideline developers create gold standards, while avoiding criticisms of standardization?*

Producing standards that regulate without standardizing

Clinical practice guidelines have been criticized for a number of different (and contradictory) reasons. The principles of Evidence Based Medicine criticize reliance on 'expert opinion' and decry the limited impact guidelines have on regulating clinicians and standardizing practice. Others denounce over-reliance on quantitative evidence, objecting to standards that ignore the

preferences and characteristics of individual patients and undermine the autonomy and expertise of physicians. In other words, literature about EBM guidelines repeats the common narratives of the promise of standards as 'rational ideals', providing universal evidence to bring practices 'up to standard', and the danger of standardization to transform the 'art' of medicine into a dehumanized industrial uniformity. The overall question I seek to answer in my dissertation is how guideline producers justify and adapt their work in response to the multitude of promises and critiques that surround EBM guideline development. Investigating how guideline developers manage the seemingly contradictory demands of guidelines that are at once based on universal evidence and sensitive to human diversity, offers an opportunity to analyze how tensions between universal and local, between 'evidence' and 'values', between standards and individuality, between science and care are managed, avoided, or reconciled in the practice of EBM.

Previous research suggests that EBM proponents have responded not by dispelling objections, but by incorporating critique into EBM's principles, for example by emphasizing the importance of "clinical expertise" and "patient values" in the *application* of evidence (Lambert, 2006). EBM's support for a diversity of knowledge in the *application* of Evidence does not alter the formal definition and hierarchy of evidence in EBM (Buetow & Kenealy, 2000), and Lambert calls at least some of the assimilationist response rhetorical strategy not borne out in reality (Lambert, 2006). Claims that a "new EBM" is developing that is "more open to the integration of different forms of evidence" thus deserve empirical scrutiny (Goldenberg, Borgerson & Bluhm, 2009:165). To gain insight in the way guideline developers achieve the challenging task of producing 'gold standards' amidst strong critique on the nefarious effects of standardization, I investigate what guideline producers consider 'good' guidelines, and what technologies, discourses or instruments are used to guarantee the

production of such guidelines. The first part of my dissertation presents how *in practice* national (in France and the Netherlands) and provincial (in Ontario) guideline developers mobilize, combine and 'add up' diverse kinds of knowledge, procedures and materials to reach collective decisions within a guideline development group. I not only provide a description of the locally situated nature of knowledge assemblage and classificatory activities (as previous research has done for such activities in both clinic and science), but this work also offers insight into the more interesting question of "how practitioners reflexively try to work around these contingencies." (Keating & Cambrosio, 2009:340). The second part of the dissertation focuses primarily on the development of meta-standards (guidelines for guideline development) that formalize and legitimize local and pragmatic guideline development, which includes the involvement of patients and their 'values' in EBM guideline development.

Methods & material

Following tenets of the field of Science & Technology Studies, the current study is not restricted to the "social" dynamics of guideline development group members, but considers them in conjunction with epistemological questions on the nature of evidence as well as 'political' questions about the regulation of medicine and professionals. Furthermore, rather than taking dichotomist categories as given, I investigate how myriad boundaries and contradictions, – i.e. between local and universal, between fact and value, between evidence and opinion – are constructed and managed in practice. It does this by not only investigating the end product of science, or what scientists say they do, but by observing "science in action" (Latour, 1987). To empirically analyze evidence-based guideline development 'in action' I gathered diverse empirical material that includes EBM's formal principles and writings (what guideline developers ought to be doing), reports,

meetings and interviews about guideline development (what they say they are doing), and observations of guideline development practices in real-time (what they are doing). I conducted intermittent (participatory) observations over a period of four years (2007-2011), consisting of attendance at two distinct types of meetings on guideline development. At the first type of meetings I observed the development of national or provincial clinical practice guidelines on a specific topic, while the second type of meetings involved the development of (international) methodologies for guideline development, or 'guidelines for guideline development'.

My observations of the development of guidelines on particular clinical topics were conducted at three institutions that all started producing guidelines in the 1990s according to the principles laid out by the new EBM 'paradigm'. Chapters three and four address how guideline developers understand and perform the nature of evidence and EBM in their practices, the first based on observations at the French program called "Standards, Options, Recommandations" (SOR), and the second set of observations was expanded to include guideline development at the Institute for Health Care Improvement (CBO) in the Netherlands, and the Program in Evidence-Based Care (PEBC) in Ontario, Canada. The SOR program was established in 1993 by the National Federation of the French Comprehensive Cancer Centers (FNCLCC), given the mandate to develop and update oncology guidelines in order to harmonize "clinical practices between cancer centers concerning diagnostic, classification, treatment and follow-up procedures" (Fédération Nationale des Centres de Lutte Contre le Cancer, 1994:50). This program relies on a distinctive framework for the production of guidelines that emphasizes the need to follow the tenets of Evidence Based Medicine by producing recommendations resting on the best available scientific evidence or on expert consensus when adequate evidence appears to be lacking (Fervers, Hardy & Philip, 2001). Between 1993 and 2006, SOR published 81 guidelines, which accounts for 54% of

the 148 clinical practice guidelines published in France (Castel, 2009). The Dutch Institute for Healthcare Improvement (CBO) “develops products, instruments, and methods for quality improvement and care innovation” which includes guidelines but also projects for improving patient flow and patient safety (CBO, 2007). CBO develops 15 to 20 evidence based guidelines per year on a very broad range of diseases and conditions, and has a well-developed Patient & Public Involvement program, including collaborations with patient advocacy organizations who want to develop guidelines. Since 1997, CBO has been collaborating with the Dutch Cochrane Center and The Dutch College of General Practitioners (NHG), leading to the establishment of methods for ‘Evidence Based Guideline Development’ (Evidence Based Richtlijn Ontwikkeling - EBRO), creating the EBRO platform in 2001. The EBRO guideline development manual is published by CBO, and was most recently updated in 2007 (EBRO, 2007). In Canada, the Program for Evidence Based Care (PEBC) has its origins in the department of Clinical Epidemiology and Biostatistics at McMaster University, which is the so-called birthplace of EBM. In 1994 the ‘Practice Guideline Initiative’ was integrated into what is now Cancer Care Ontario, and a ‘systematic’ methodology called the ‘practice guidelines development cycle’ was devised (Browman et al., 1995). PEBC is a member of the Cancer Guidelines Action Group of the Canadian Partnership Against Cancer that has established the ‘Inventory of Cancer Guidelines’ which is an electronic database that houses cancer guidelines from all over the world, and which have been evaluated and rated for quality. Interestingly, the PEBC organization included lay involvement in guideline production in the early years of PEBC, but this involvement has almost completely disappeared.

Although SOR and PEBC exclusively develop cancer guidelines and CBO includes a broad range of topics, I consider them together as ‘EBM guideline developers’ since the principles and process of guideline development (rather than the clinical topics addressed) are similar, albeit

with some institutionally specific arrangements. The institutions and people I focus on in this study develop guidelines (and related quality control instruments, such as performance indicators) as their main activity. The employees at CBO, PEBC or SOR conduct the literature search, extract evidence tables and write large parts of the guideline text. They also invite, organize and chair guideline development groups (GDGs), for each topic a unique group of 5-20 experts is set up, including academics (for their research expertise), practicing clinicians (for their clinical expertise) and sometimes patients (for their 'values' and/or experiential expertise). While the experts on each GDG of course are guideline developers when they participate in the GDG, I refer to them as GDG participants, to distinguish them from the experts in guideline development (and its methodology), sometimes referred to as 'methodologists'. This dissertation focuses primarily on the discourses, practices and instruments that these experts in guideline development use and produce in order to develop 'good' guidelines.

The social organization of evidence-based decision making and tools may differ depending on the specific health care systems in which they occur (Mykhalovskiy & Weir, 2004:1062), and yet the guidelines of all three institutions are part of health care systems that have similar arrangements that combine professional regulation with public management and financing (especially compared to the fragmented and predominantly privately funded and managed system in the USA, where guideline development (and EBM in general) seems to have taken a different turn). The guidelines produced by SOR, CBO and PEBC are neither legally nor financially binding, but they are endorsed by professional and public bodies who use them as references in the development of other standards and instruments, such as education, performance indicators, audits or reimbursement policies. These guideline organizations each produce about a dozen guidelines per year that are highly regarded in their respective province, countries and even

internationally. Significantly, researchers and guideline developers at these organizations train other guideline developers in evidence based guideline development through national and international workshops, meetings and summer institutes for novel guideline developers. The (former) directors and guideline developers of CBO, PEBC and SOR (e.g. Jako Burgers, Kitty Rosenbrand, Melissa Brouwers, Beatrice Fervers) publish on guideline methodology, including articles co-authored together with other (inter)national guideline developers and researchers. They were involved in the founding of the AGREE collaboration that calls its mission to "advanc[e] the science of practice guidelines"[1] and that directly led to the establishment of the Guidelines International Network (GIN) in 2002, and of which CBO, PEBC and SOR are members. As such, the people, principles and practices at these sites have been crucial in the establishment of international standards for guideline development. And while the guideline practices of SOR, CBO and PEBC may have institutionally distinct features, they are comparable since their international collaboration informs and harmonizes their evidence based methodological principles and practices.

The activities of The Guidelines International Network (GIN) provided the focus of my second set of (participant) observations, as they are at the center of international regulation and standardization of guideline development. GIN aims to "improv[e] the efficiency and effectiveness of evidence-based guideline development, adaptation, dissemination and implementation"[2] and has developed international standards for guideline development (Qaseem et al., 2012). Every year, about 400 guideline developers from dozens of countries around the world gather at GIN's annual conference to present the latest developments on the "science of guideline development". They share practices, problems and solutions, set

---

[1] www.agreetrust.org/ , accessed December 4th 2012.

[2] www.g-i-n.net/about-g-i-n , accessed July 15 th 2011.

up working groups and endorse and promote a host of other instruments (AGREE, GRADE, ADAPTE and CoCanCPG). In 2007 GIN established a working group on patient and public involvement in guideline development, called G-I-N PUBLIC, which I joined as a participant observer.

The empirical material collected at these sites includes medical literature, manuals, documents, instruments and presentations used and produced by guideline developers to guide the EBM guideline developmental practices. I participated in and observed training and networking events organized or attended by CBO and PEBC guideline developers, including five subsequent annual meetings of the Guidelines International Network; two summer institutes (one on 'knowledge translation' and one on guideline development); and one team meeting at PEBC. I worked as a guideline reviewer with AGREE, and from 2008-2010 I acted as participant observer at GIN's working group on Patient and Public Involvement (G-I-N PUBLIC), initially I participated as website moderator, then as steering committee member, and after data collection was completed in 2010 I acted as co-chair. Detailed fieldnotes were taken at all these events and meetings. From September 2008 until November 2010 I attended meetings of various guideline development groups (GDGs), ensuring the inclusion of a broad range of guideline topics at different stages of development. I observed one meeting at SOR in France, fourteen meetings at the CBO in the Netherlands (on six different guideline topics) and eight teleconference meetings at the PEBC in Canada (on a single guideline for a specific condition). All meetings were audio-recorded, and the SOR meeting transcribed in full, for the subsequent meetings I relied on detailed fieldnotes to transcribe specific passages of interest. The notations used refer to the guideline development group (GDG1, GDG2) and the sequential number of the meeting I attended (a,b,c).

I conducted a total of 22 semi-structured, hour-long interviews with key figures working in various (overlapping) capacities within the world of

guideline development: including 'methodologists' responsible for developing specific guidelines at CBO, PEBC and SOR; members and staff of patient organizations involved with guideline development (denoted as *PtRep1-5*); and the program manager of public funding initiative for innovative guideline development at the Netherlands Organisation for Health Research and Development (ZonMW) that funds (with 'strings attached') many guidelines in the Netherlands. Interviewees also included researchers actively involved in developing new methodologies for guideline development, such as founding members of the Guidelines International Network who are involved in international collaborations such as AGREE and ADAPTE, and the principal organizer of the GRADE collaboration, who coined the term Evidence Based Medicine back in 1991. If relevant, the names of interviewees or their organization are reported, since the visibility and uniqueness of the work of interviewees or their organization made anonymity not feasible (or desirable). Interviews were recorded and transcribed in full. To maintain confidentiality of GDG participants, certain information in direct quotes taken from GDG meetings was altered (such as the guidelines' topic, as well as the names of specific treatments or authors). I translated material that was not in English. The study was approved by the Research Ethics Board I of McGill University.

Organization of the dissertation

Chapter two provides a detailed review of the literature on EBM, medical standards and guideline production from which this dissertation draws. The empirical findings are presented in four subsequent chapters, which are divided in two parts.

The first part consists of chapters three and four that analyze how the nature of Evidence and EBM is understood and performed in the practice of producing specific guidelines. They are based on observations of the

meetings of guideline panels that produce guidelines at a French, Dutch and Canadian institution, complemented with in-depth interviews with these guideline developers. It analyzes the production of guideline texts in the material sense of the term, i.e., as a set of sentences and paragraphs formulated by the multidisciplinary panel. In chapter three Foucault's notion of 'dispositif' is invoked to demonstrate that, while evidence is important, it is but one element of institutionally specific procedures, rules and distinctions that make the production of a guideline text possible. Considerations external to the guideline panel (such as existing practice and peer review) are important elements of these procedures. Chapter four further expands this analysis by specifically focusing on the procedures that guideline developers employ to define evidence as a category (regardless of its level) and to distinguish it from other kinds of knowledge. The term "evidence searched guidelines" is introduced, as my analysis shows that the defining feature of EBM guidelines is not the quantity or quality of evidence on which the guideline is based, but rather the methodological process of evidence searching, selection and presentation. Various kinds of 'non-evidence' are legitimately relied upon at all stages of EBM guideline development. Increasingly procedures exist to manage 'non-evidence', challenging the accounts of EBM's proponents and critics alike.

The second part consists of chapters five and six and investigates discourses and instruments that seek to guarantee the quality of guidelines, particularly the 'guidelines for guideline development' of the Guidelines International Network, and methods of Patient & Public Involvement that aim to make guidelines (or their production) more democratic, of higher quality, more useful, more objective. The empirical material included here is based on participatory observations conducted at workshops, conferences and research projects on the production, evaluation and regulation of guidelines, as well as interviews with guideline developers and researchers who developed widely used methodological instruments, including those involved

in the practice and methods of patient and public involvement. Chapter five focuses on the Guidelines International Network (GIN), and I argue its distinct meta-standards are indicative of a new 'institution of meta-regulation' in medicine (Cambrosio et al, 2006). Established in 2002, with funding and support from the European Union, GIN brings together numerous organizations around the world that produce thousands of (conflicting) guidelines. Despite EU's emphasis on harmonization within Europe, GIN eschews 'Europeanization' of national guidelines, emphasizing the importance of 'organizational context' and 'cultural values' in standard setting. Instead, GIN develops 'guidelines for guideline development' that standardize the guideline development process. This shift from product standardization to process standardization allows GIN to legitimize duplicate guidelines by making their development more universal, transparent and regulated. I argue this helps legitimize the work of the nascent and much criticized 'profession' of guideline methodologists. Furthermore, I argue that the regulation of guideline developmental process present a novel attempt to address the contested relation between the individualized (or local) and the standardized (or universal) within the EBM paradigm. Chapter six analyzes a different effort to legitimize, regulate and improve guidelines: the involvement of patients and their 'values' in EBM guideline development. ? This chapter provides a review of the literature on PPI in health care and medical science, and draws on empirical material from participant observation, interviews and document analysis of those involved in the promotion, planning and practice of PPI in guideline development. As a member of the Guidelines International Network working group on patient and public involvement (G-I-N PUBLIC), and thus participant as much as observer, my aim is not limited to contributing to the literature on PPI. By addressing objections to and limitations of PPI models, increasing our understanding of the distinct rationales and by highlighting the benefits of diversity in PPI models and methods, I hope this chapter can assist those

looking for guidance on putting PPI in practice. The chapter describes four distinct rationales for 'lay' participation. The first and second model reiterate the 'consumerist' and 'democratic' models of existing literature on patient and public involvement in health care, drawing respectively on consumers' right to be involved in choosing their own and personalized care and the 'social' rights of citizens and taxpayers to base policy on public values. The third model, Lay Expertise, emphasizes patients' experiential knowledge to contribute to 'context-sensitive' guidelines. In the forth model I introduce a novel rationale for PPI in guideline development, as 'Critical Witness' they can provide epistemological oversight that strengthens the evidence based approach. Finally, I argue that diversity in PPI is not an obstacle to establishing PPI as a 'universal standard', but an essential feature to achieve PPI.

In conclusion, chapter seven briefly summarizing the main findings and provides a synthesis of how the findings answer the overall research question I started out with, and situate these findings within existing debates about the nature of Evidence in EBM and the regulatory consequences of clinical practice guidelines. For an introduction into this literature, the next chapter presents a review of the existing debates on EBM and its guidelines.

CHAPTER 2 Literature Review

Debating evidence, regulation and patients in guidelines

The emergence of evidence based guidelines

To provide a short history of how clinical practice guidelines (CPGs) emerged out of a variety of attempts to improve the quality of health care, I draw on the work by Weisz and colleagues (Weisz et al, 2007). While the standardization of terminology, disease classifications and clinical *research* has a much longer history, it is not until the 1980s that standards for *clinical practice* - documents that 'tell doctors what to do' – start to be produced on a large scale. Design and terminological standards had made disease categories and medical records comparable across institutions, states and countries, making large scale and comparative medical research possible, such as randomized clinical trials and epidemiological research. In the USA during the 1970s – thanks to the introduction of Medicare and Medicaid – medical treatments provided by a great number of doctors started to be registered and tracked, and these patterns showed considerable variations in practice across regions (Wennberg & Gittelsohn, 1973). Combined with concerns over the effectiveness and rising cost of common health care interventions, the question was raised as to which of these diverse practices presents higher quality and/or greater efficiency? It was under these circumstances that providing individual clinicians with knowledge of the "best practice" to follow (i.e. a gold standard) was considered an important tool to raise the standard of medical care. Many different organizations in countries all around the world start developing clinical practice guidelines (CPGs), recommending a standard course of action, a certain diagnostic test or a

suggested course of treatment for the medical condition in question. The NIH and other institutions started consensus conferences in the 1970s and 1980s to establish practice standards based on the expertise of national experts (Jacoby, 1993), and in 1990 the Institute of Medicine established the definition of Clinical Practice Guidelines as “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances” (Field & Lohr, 1990:8). This remains the most quoted definition and was only revised for the first time in 2011). When evidence based medicine was pronounced a new ‘paradigm’ for medicine in 1992, it brought renewed interest in guidelines, modifying guideline development methodology to ensure gold standards for practice were evidence based.

The term Evidence Based Medicine was coined in 1992 by Gordon Guyatt, David Sackett and their colleagues at McMaster University in Hamilton, Canada who pronounced EBM as a ‘paradigm’ shift in medicine, away from “intuition, clinical experience, and pathophysiologic rationale” and instead “using the literature more effectively in guiding medical practice”; it was to be a more scientific basis for medicine (Evidence Based Medicine Working Group, 1992: 2420). Distrust of intuition and unsystematic clinical experience applied not only to local practitioners, clinical advice of national experts was also relegated to “the way of the past” (idem:2420). Clinical practice guidelines were “only worth following” when they were based on an exhaustive review of the medical literature, critically appraised and assigned a ‘level of evidence’ (Sackett & Rosenberg, 1995:623). Meta-analysis and randomized clinical trials (RCTs) are categorized as the highest level of evidence, and case reports or expert opinion are at the bottom of the evidence hierarchy (see figure 1). As it became clear that individual clinicians lack the time and skills to critically assess medical literature themselves, EBM increasingly advocates reliance on ‘evidence reports’ that synthesize, summarize and review evidence on a specific topic for clinicians,

such as systematic reviews produced by the Cochrane Collaboration, and guidelines that rely on such evidence reviews to justify recommendations (Guyatt et al, 2000). As of 2011 the Institute of Medicine has now modified its definition of guidelines according, specifying guidelines are to be "informed by a systematic review of evidence" (Institute of Medicine, 2011).

Since its emergence, EBM has become an important approach to contemporary western medicine and is taught in medical education, strived for in health care policy and praised and celebrated in editorials in medical journals. The principles of EBM have been so widely applied and supported that we now have 'evidence-based everything', including nursing, policy and bioethics (Fowler, 1997). Nevertheless, EBM has been the target of much critique by medical professionals, epidemiologists, philosophers and social scientists (Lambert, 2006; Mykhalovskiy & Weir, 2004). It has been characterized as a 'social movement' and a 'reform movement' as it is the topic of myriad polarizing debates about how medicine ought to be practiced and regulated (Pope, 2003). Here I highlight three central themes of the debates around EBM: the way EBM defines the nature of knowledge; the regulatory impact that EBM's standards have on the medical profession; and the role of patients in 'good' medicine.

Debating the nature of knowledge: the politics of evidence

The Evidence Hierarchy (figure 1), by (re)defining the kind of knowledge needed to provide a basis for good (rational) medicine, is the novel and 'paradigmatic' characteristic of EBM. To distinguish this EBM-specific characterization of evidence from the common notion of evidence that includes more diverse kinds of knowledge, proofs and justifications, I will refer to the former as Evidence (with capital E). The categorization of medical knowledge in a hierarchy of evidence has been a main target of

critique, by medical practitioners, philosophers and social scientists (Lambert, 2006; Mykhalovskiy & Weir, 2004), and it has been identified as "the guilty source of [EBM's] questionable epistemic practices" (Goldenberg, 2009: 171). Critiques primarily lament the priority given to the quantitative evidence produced by randomized controlled trials (RCTs), and RCTs' limits, flaws and faults are discussed in detail (Grossman & Mackenzie, 2005; Cartwright, 2007; Bluhm, 2005; Borgerson, 2009). One of the main problems with RCTs is that their results are difficult to apply to 'real' patients. In order to achieve high internal validity (i.e. it is likely that the results are accurate for the group included), RCTs sacrifice external validity, i.e. the outcomes of the selected research subjects may not be valid for 'real' patients typically seen in the clinic who differ in many relevant ways from the study population (age, sex, co-morbidity, undergoing multiple treatments). For example, the validity of RCT evidence has been questioned because experimental subjects are typically younger and healthier, and more often white and male, than patients in the clinic (Epstein, 2007). So while RCTs are valuable to answer certain questions, they are not a universal gold standard as in many situations RCTs are unnecessary, inappropriate, impossible or inadequate, and, for example, observational studies - based on patient populations from regular clinical practice, rather than highly selective research subjects – may be more informative and relevant (Black, 1996). Moreover, on the basis of methodological design alone, the evidence hierarchy excludes entire scientific disciplines, and neglects more complex social, political or public health interventions that cannot be (easily) tested in RCTs. This includes "behavioural, psychosocial, community-based and multiple-component interventions" (Lambert, 2006:2635); complementary and alternative medicine (Borgerson, 2005); studies on the determinants of health and public health interventions (McGuire, 2005), as well as medical anthropology, qualitative research and other 'narrative-based' knowledges (Lambert, 2006). Moreover, since RCTs are very costly and chiefly

conducted by the pharmaceutical industry, EBM "gives financial interests unprecedented power to shape medical practice." (De Vries & Lemmens, 2006:2695), as industry-sponsored research presents serious problems such as "selective publication of trial data and ghostwriting of publications" (Healy, 2009:16). Goodman thus concludes that "even our best tools for reducing bias fail to eliminate it; and some of these tools even present new sources of bias." (Goodman, 2003:43). Goldenberg concurs, reminding us that theories (and actions) are underdetermined by evidence such that RCT protocols can never eliminate "culture, contexts, and the subjects of knowledge production from consideration", but rather they aim to "obscur[e] the subjective elements that inescapably enter all forms of human inquiry" (Goldenberg, 2006:2622). These kinds of writing all share the message that evidence can not replace or erase politics, and draw attention to the politics of evidence itself, as EBM's (re)definition of evidence in- and excludes various kinds of knowledge and knowers.

Mykhalovskiy and Weir suggest new directions for social science research into the politics of Evidence because the literature in this area is "empirically thin and pitched at high levels of abstraction" (2004:1061). This dissertation responds to their call for more empirical research of EBM and its instruments in practice. Chapters three and four examine the critical question 'what counts as Evidence', but unlike existing abstract and normative critiques of the formal evidence hierarchy, I investigated how Evidence - and the lack thereof - is defined, categorized, included and excluded during the actual production of clinical practice guidelines. Like work on the pragmatic achievement of (diagnostic) classification in clinical practice (Latimer et al., 2006; Bourret & Rabeharisoa, 2008; Hedgecoe, 2002; 2003), this dissertation shows evidence classification is not merely an automated 'application' of formal criteria, but requires a situated process that draws on diverse sources of knowledge and rules, and in which both classification and cases to be classified shift over time. To understand what

counts as Evidence I conducted an empirical investigation of Evidence as an emergent category produced during the practice of guideline development.

Some empirical social science work on the production of guidelines does exist, but it rarely addresses the issue of evidence and knowledge, limiting analysis to professional power relations. Using literature review, questionnaires on imagined scenarios or retrospective analysis (documents, interviews) such studies see professional status as a source of power, which can then determine knowledge, without addressing the *ways* in which external status is reproduced and inscribed in what is supposed to be a 'technical' document (Leape, Park, Kahan & Brook, 1992; Kahan et al., 1996; Pagliari & Grimshaw, 2002; Raine et al., 2004; Hutchings & Raine, 2006;). However, knowledge is not simply the result of professional politics, rather it is also a *source* of power that can be used to claim authority. Likewise, guidelines do not simply reproduce existing professional relations but they also determine and change professional status, authority and work (Castel & Merle, 2002; Zitzelsberger, Grunfeld & Graham, 2004; Berg, Horstman et al., 2000).

A few studies have conducted retrospective reviews of (conflicting) guidelines and/or interviews with guideline developers to analyze the kinds of knowledge and rationalities mobilized to manage the contested relation between ‘local’ context and ‘universal’ science (Dobrow et al., 2006; Will, 2009; Berg, Ter Meulen & van den Burg, 2001). Dobrow and colleagues argue that guideline development is dominated by “rule based decision making” (evidence hierarchies) and not “agreement based decision making” (stakeholders’ input) and concludes there is an over-focus on the quality of evidence (level of evidence), while very limited attention is paid to the applicability of evidence to the ‘context’ and patients outside the trial setting. Thus concluding that the limitations of the Evidence hierarchy are reproduced in guideline development (Dobrow et al, 2006). Berg and colleagues assessed the “normative assumptions” about how guidelines can

contribute to ‘good’ medicine and ‘rational’ decision making. They distinguish two kinds of rationality: “formal rationality”, typically used by methodologists and specialists and “practical rationality”, more often used by (general) practitioners. They also suggest that formal rationality should inform the more pragmatic rationality, but should not replace it. Will (2009) analyzed two divergent guidelines on the same topic and shows that the extent of the applicability of RCT evidence depends on whether trial methodology is interpreted as proof of protocol or proof of principle. If trials are considered to provide proof of protocol, tools and decision based on such evidence must closely resemble (or undertake to mimic) the standardized environment of the trial protocol. If guideline developers consider trials to provide proof of principle, they rely on biological and ethical principles to ‘translate’ the trial evidence to less standardized practices and to patients outside the trial population. In chapter four, I extend Will’s findings by adding ‘current clinical standards’ as a third type of knowledge relied upon to *apply* or ‘generalize’ trial evidence. Moreover, chapter four demonstrates principles and ‘local’ knowledge are also relied upon in determining what counts as Evidence in the first place.

The above studies rely on analysis of the final document or formal reporting of the guideline process; in other words they were limited to the studying the end product of science, or to what scientists *say* they do. The observation of “science in action” can provide new insights into what knowledge production is like (Latour, 1987). To analyze what guideline makers actually do to arrive at the end product, I thus chose to observe guideline development group (GDG) meetings in real time. To my knowledge, only one other study has studied guideline production ‘in action’ (Moreira, 2005, Moreira et al, 2006). Extending Moreira’s approach, this study recognizes the importance of evidence, while considering it but one element of the justifications, rules and distinctions that make the

production of CPGs possible. I extend his work by focusing on the material nature of guideline development as the production of a *text* (chapter three).

Invoking the notion of 'dispositif' to emphasize the importance of institutionally specific rules and procedures to structure the GDG debates. Chapter four furthermore argues for a specific interrelation of the types of justification that Moreira (2005) claims guideline developers rely upon: robustness (science), acceptability (politics), usability (practice) and methodological adequacy (process). I propose the term 'evidence searched guidelines' to indicate that the EBM-ness of guidelines does not depend on its evidence-base, but on a specific 'process' repertoire on how to search, select, assess and present evidence. I argue the methodological adequacy repertoire in fact functions as a 'meta-repertoire': it provides rules on how to construct, mobilize and present the science, politics or practice repertoires. Chapter five describes the emergence and consequences of universal methodological standards for guideline development, such as AGREE and GRADE.

Debating the regulatory power of standards

The EBM movement emphasizes the purpose of evidence-based guidelines is to rationalize medicine, and they differ from other educational tools (such as textbooks) in that they do not simply provide a catalogue of good medical treatments, but have direct and explicit regulatory aims, being produced "with a definite intent to influence what clinicians do" (Hayward et al., 1995). Much social science work has examined clinical practice guidelines as regulatory tools, investigating their impact on medical professionals' autonomy and authority and shifting power relations within medicine. By specifying 'good practice', guidelines open "the black box of clinical judgment" providing norms and quality standards to external parties (managers, government, insurers, patients) who can evaluate the

professionals' behavior and hold them accountable (or deny payment) for other than (gold) standard treatment (Timmermans, 2005:499). As guidelines may "serve as a foundation for instruments to evaluate practitioner and health system performance" (Field & Lohr, 1990:2-3), the medical profession might not only lose the monopoly to perform medical work, but also the privileged authority to monitor, evaluate and judge that work, another marker of professionalism (Freidson, 1970). Critics of guidelines have invoked the term "cookbook medicine" to express the danger of guidelines to deprofessionalize doctors, reducing medicine from a clinical art to simply following a 'recipe' in which clinical skills and expertise are ignored (Lambert, 2006). Such accounts reiterate a more general argument of reduced trust in professionals' self-regulation and autonomy, so that in the late 20th century external parties such as managers and government increasingly employ evaluative tools to monitor, audit and regulate professionals (Power, 1999). Professionals' critique on guidelines are characterized as resistance to (or 'decoupling' of) external auditing measures, in an effort to preserve their professional power, autonomy, skills and trust (Power, 1999; McGivern & Fischer, 2012). Many of the accounts on the regulatory impact of guidelines share an assumption that guidelines are imposed externally and resisted internally, and conclude, for better or worse, that guidelines reduce the autonomy and authority of medical professionals. Yet, EBM's pervasive "technologies for transparency" (Blomgren & Sahlin, 2007: 161) are produced and managed by (or in participation with) medical professionals and their organizations rather than imposed by external parties (Levay & Waks, 2009; Weisz et al., 2007). The profession's enthusiasm for, rather than resistance to, EBM's standards has been interpreted as a defensive strategy to avoid externally controlled regulation by providing their own quality control that is controlled internally by the profession. Instead of third parties it is the 'knowledge and administrative' elite of the medical profession that employs EBM's standards "to order, assess and direct the work of the

rank and file" doctors, creating a new kind of hierarchical intra-professional regulation (Freidson, 1984: 15-16; Armstrong, 2002). In this account the collective autonomy of the medical profession has increased at the cost of diminished individual autonomy of 'rank and file' physicians (Freidson, 1984). While this shift from 'external' to 'intra-professional' control redraws the boundaries of who constitutes the regulators and the regulated, it continues to assume only two parties are important: those who monitor, measure, see and control; and those who are being monitored, measured, seen and controlled. And the relation between them is assumed to be a one-way hierarchy: standards and transparency are a threat to the professional autonomy of those whose work is audited/standardized. A crucial element is missing in such accounts: *what* is being standardized or made visible (how and by whom) matters. Empirical studies of "how guidelines actually affect health care", and others that examine the *kinds* of visibility and standardization produced by guidelines, tell a more diverse story. This study contributes to such empirical studies by focusing on the production and producers of a specific standard – clinical practice guidelines, as well as on the regulation and regulators of guidelines.

Empirical studies of the use of standards have emphasized that standards that make professionals' work more visible can both increase and reduce professional authority and autonomy. Especially for weaker professional groups or individual professionals, increased visibility of their work and skills can enhance the scientific status and legitimacy of their work (Levay & Waks, 2009: 523; Timmermans and Berg, 2003). Professional standards can help claim or defend a unique domain of expertise (Bowker and Star 2000; Berg, Horstman et al., 2000) or challenge existing professional hierarchies by changing intra-professional competition and communication (Castel and Merle 2002; Castel, 2009). To function as a professional *resource* rather than an external *restraint,* standards reflect and explicate existing work practices (rather than modify them) and build in

flexibility and ambiguity that requires professional judgment and skills to understand and manage (Berg, Horstman et al, 2000: 773-5). When designing their own standards, professionals remain in control of *what* is made measurable, comparable and governable, thereby preserving the need for professionals' autonomy and authority in performing and evaluating practice. Central to the inquiry into the regulatory impact of standards must be the question 'Who controls the quality control tools?'. This implies not only inquiry into the *use* of evaluation tools (Power, 2000), but also inquiry into who controls the *production* (and design) of such tools, to understand what is being evaluated in the first place. Chapters five and six of this study examine the regulation of guidelines and their production, examining the tools, standards and discourses that control the quality (and legitimacy) of guideline development. Chapter five presents the Guidelines International Network as new center of meta-regulation within medicine, while chapter six introduces four models of patient and public involvement for improved quality of guidelines.

Empirical studies of the use of standards have also challenged the rhetoric of standardization simply replacing the local, particular or art of medicine with a universal rationality. Rather than seeing the universal, standard or science as opposite to – and aiming to replace – the 'disorder' of practice, 'order' can only come into being by incorporating 'disorder', and both standard and practice are modified in the process (Berg, 1998). These authors "stress the multiplicity, joint emergence, and the symbiotic relationships between the orders and their disorders." (Berg & Timmermans. 2000: 59), and emphasize that "[w]hen the 'local' is seen as always part and parcel of everything universal, the latter need no longer be perceived as a necessary (latent) threat." (idem:60). They show in a range of settings that guidelines and protocols do not turn workplaces into the smoothly working "well-oiled machines" of which proponents of evidence based medicine dream, nor into the "stifled robot-scapes" or 'cookbook medicine' that critics

of guidelines fear (Timmermans & Berg, 2003: 56). CPGs cannot invent a clinical practice de novo[3], but need to take hold in an existing practice of existing routines and "networks already firmly in place", standards being but "one more element that was inserted into a set of already existing interests, associations, and practices" (Timmermans & Berg, 1997: 274). Standards can not function if they are not embedded into the existing (net)work, and can only be ignored. In order for standards to function, much work is needed, not only to adapt the local context to the standard, but also to adapt the standards to the specific context of the practice: "In trying to get the 'order' of the protocol to actually *work,* thus, it inevitably becomes swamped with more and more disorder. […] a new order is achieved that *incorporates* the very messiness it started out to erase" (Berg, 1998: 238,246, emphasis in original). Even if standards are designed to change and replace an existing practice and routine, "the same standards need, to a certain degree, to incorporate and extend those routines" (Timmermans & Berg, 1997: 274). The authors thus coin the term 'local universality' to describe successful standardization in practice.

Not only critics of or commentators on EBM acknowledge the importance of 'the local context'. Proponents and practitioners of EBM have equally become aware that "[t]he use of evidence is most successful when local differences are factored into the decision-making process", promoting efforts to "globalize the evidence, localize the decision" (Eisenberg, 2002). While this acknowledges that more than evidence is important in the practice of EBM, the notion of universality of evidence itself is maintained, and it is unclear whether and how standards can mediate and connect such distinctly different worlds of global evidence and local decisions. Are the local

---

[3] As an exception, regulatory standards may raise their own networks and practices de novo. In the field of clinical cancer genetics, standards and regulatory conventions determine to a great extent what happens "in the privacy of the consulting room". This may be because "professional regulatory interventions occurred as the field emerged, rather than, as is usually the case, after the novel practices reach a mature stage" (Bourret, 2005:64).

adaptations that Berg & Timmermans outlined as necessary for successful standardization also mobilized in the production of guidelines, creating more localized standards, as has been suggested in case of more local clinical pathways (Allen, 2009) and institutional procedures (Nes & Moen, 2010). Guidelines are but one element along this chain of textual translations linking knowledge production to the regulation of medical practices (Cambrosio, Keating, Schlich & Weisz, 2006). As an instrument of 'regulatory objectivity' guidelines regulate, organize and connect medical science and clinical practice, simultaneously presenting a new kind of knowing (science, evidence) as well as a new kind of regulation (standard, control), and this study investigates how the principles and practices of guideline developers manage tensions between science and care; quantitative evidence and human skills; standards and diversity; universality and local specificity. One such specific (apparent) dichotomy is addressed in the last chapter, that of evidence based versus patient centered.

Debating the role of the patient in EBM

Evidence-based guidelines have been criticized for being "doctor-centered" and "erasing the patient" (Mykhalovskiy & Weir, 2004). Bensing states that because EBM regards medicine as a "cognitive-rational enterprise", "the uniqueness of patients, their individual needs and preferences, and their emotional status are easily neglected as relevant factors in decision-making." (2000:17). The rationalizing and standardizing ambitions of EBM have led to the charge that "EBM undermines humanism in medicine by discounting the individual patient, viewing him as a mere population statistic" (Brody, Miller & Bogdan-Lovis, 2005: 571). EBM's gold standards that prescribe 'the' best treatment on the basis of quantitative evidence ignore *legitimate* variation in treatment due to the preferences, characteristics, needs and experiences of individual patients (Greer et al.,

2002), and reduces patient choice and autonomy by determining choices by a guideline panel instead of the doctor-patient dyad in the clinic (Barratt 2008; Rogers, 2002). Brody, Miller and Bogdan-Lovis challenge the notion that EBM's reform makes medicine less patient-centered, instead they argue that EBM is characterized by a *different* representation of patients, not by 'less' or 'more' presence of patients: shifting away from patho-physiology - in which 'individual patients' figure prominently as organs, tissues, lab results and molecules - towards the study of populations - in which a patient is characterized by his/her age, geography, risk, gender, and behavior (smoking, diet, exercise, work). Moreover, patients and their organizations are often advocates of EBM, promoting the use of Evidence and (patient versions of) clinical practice guidelines, and EBM organizations and initiatives such as the Cochrane Collaboration and guideline development increasingly involve patients and promote shared decision making (tools) (Brody et al., 2005). The incorporation of patient preferences has been called "the next step in guideline development" (Krahn & Naglie, 2008), and although so far there are very few empirical studies available on how to do this successfully (Van de Bovenkamp & Trappenburg, 2009), the inclusion of patient values and preferences in evidence-based guidelines is on the agenda of many guideline developers and researchers. For example, in 2007 an international working group of researchers, guideline developers and patient representatives was founded to address the issue of patient involvement in guideline development (Boivin et al., 2010). Moreira describes how in the UK, despite various procedures for patient involvement in standard-setting, patients and their organizations mobilize against (certain) standards, objecting to the way 'values' had been collectivized by UK guideline developers (Moreira, 2012). Rather than answering whether EBM's initiatives to include 'patient values' (as well as 'other' knowledge such as clinical expertise) makes guidelines more or less patient centered, I follow the call by Mykhalovskiy and Weir (2004) to investigate the creation of

new patients, and the new relations between knowledge, practice, patients and practitioners that EBM's current guideline development brings about. Chapter six, by presenting four models of the contributions of patient involvement to evidence-based guideline development does not answer the provoking question whether evidence-based medicine limits or facilitates patient choice (Rogers, 2002), but does analyze what *kinds* of choices and what *kinds* of patients are supported, emphasized or ignored by EBM guidelines. Overall, this dissertation aims to decipher “what patient, what notion of medical work, what objectivity, what configuration of professionals, third parties, regulators, and so forth is constituted by a specific standard” (Timmermans & Berg, 2003:200), specifically clinical practice guidelines produced according to EBM principles. In order to start shedding some light on these issues, chapter three will provide an analysis of the production of a specific oncology guideline in practice, demonstrating how evidence is classified, and highlighting the procedural rules and emergent debate that make the production of a guideline text possible.

CHAPTER 3

Pragmatic evidence and textual arrangements: a case study of French clinical cancer guidelines

Authors: Loes Knaapen, Hervé Cazeneuve, Alberto Cambrosio, Patrick Castel and Béatrice Fervers

Authors' contributions : BF, HC and PC conceived of and designed the study, LK & HC gathered data, all authors advised on interpretation of data, LK and AC conducted primary data analysis, LK wrote the first draft of the manuscript, which PC and AC critically revised it for important intellectual content, all authors reviewed and approved final manuscript.

This chapter first appeared in: *Social Science & Medicine* (2010) 71: 685-92

Abstract

Both critics and supporters of Evidence-Based Medicine view Clinical Practice Guidelines as an important component of this self-defined "new paradigm" whose goal is to rationalize medicine by grounding clinical decision-making in a careful assessment of the medical literature. We present an analysis of the debates within a Guideline Development Group (GDG) that lead to the drafting, revision and publication of a French cancer guideline. Our ethnographic approach focuses on the various aspects of the *dispositif* (or apparatus) that defines the nature and roles of participants, procedures, topics and resources within the GDG. Debates between GDG members are framed (but not dictated) by procedural and methodological rules as well as by the reflexive critical contributions of the GDG members

themselves, who justify their (tentative) recommendations by relating to its (possible or intended) audiences. Guideline production work cannot be reduced to an exchange of arguments and to consensus-seeking between pre-defined professional interests. It is about the
production of a text in the material sense of the term, i.e., as a set of sentences, paragraphs, statements and formulations that GDG members constantly readjust and rearrange until closure is met. As such, guidelines partake in the emergence and stabilization of a new configuration of biomedical knowledge and practices grounded in the establishment of mutually constitutive links between two processes: on the one hand, the re-formatting of clinical trials into a device for producing carefully monitored evidence statements targeting specific populations and clinical indications and, on the other hand, the increasingly pervasive role of regulatory processes.

Introduction

Clinical Practice Guidelines (CPGs) are key components of Evidence Based Medicine (EBM), the self-styled “new paradigm” which “de-emphasizes intuition, unsystematic clinical experience, and pathophysiologic rationale as sufficient grounds for clinical decision making”, and argues for “using the medical literature [especially results from randomized clinical trials] more effectively in guiding medical practice” (EBM Working Group, 1992: 2420). Medical reformers and administrators consider CPGs “the tool of choice to weed out unwarranted variation in diagnostic or therapeutic practice and to enhance the scientific nature of medical care delivered” (Berg, Horstman, Plass, & van Heusden, 2000: 766). Thousands of CPGs have been produced in the past decades by a great variety of institutions and associations in many different countries (Weisz et al., 2007).

Unsurprisingly, CPGs have attracted the attention of many commentators. A substantial part of the social science literature on this topic (e.g., Timmermans & Berg, 2003; Castel & Merle, 2002) focuses on the *use* of guidelines as distinct from their *production*. This distinction, however, may be challenged, for, as we will see, guideline producers often openly discuss its potential use and users; as noted, more in general, by science studies scholars (Akrich, 1992), technical devices contain built-in *scripts* of their expected deployment: examining those scripts can deepen our understanding of future uses. Most articles on the production of guidelines have been published in medical journals and usually consist of methodological recommendations and suggestions on how to improve the process (e.g., Eddy, 1990; Eccles et al., 1996). Among the few social science studies that investigate guideline production, a number have resorted to an experimental or a retrospective design to correlate the professional characteristics of guideline group members with their decisions (e.g. Hutchings & Raine, 2006). Yet, as analysts of procedural rationality would argue (e.g. Reynaud & Richebé, 2007: 8), guideline development cannot be equated simply to a decision about a preset number of choices, but often leads to novel, unexpected solutions. Only ethnographic investigations of guideline development can account for the dynamics and peculiarities of processes that take place in time.

Pagliari and Grimshaw (2002) observed interactions among group members, focusing on the effect of professional role and status on group discussions. However, by implying that decisions were constrained or even pre-determined by pre-existing social variables such as professional status, their study foreclosed any consideration of their emergent nature as predicated upon interactions between group members. In contrast, Moreira's (2005) ethnographic study of guideline development more subtly portrays the debates taking place during group meetings. Borrowing from Boltanski and Thévenot's (1999) "pragmatic sociology", Moreira focused on the

actors' own critical capacity, identifying five types of "repertoires" used by the participants to justify the guideline's content by reference to the *actions* to which the guideline would presumably lead in the "external world". While participants from different professional groups made preferential use of specific repertoires, Moreira attributed this fact less to the presence of *a priori* interests than to the observation that group members envisioned different (future) practices and users.

Although there is considerable methodological overlap between Moreira's approach and ours, there are also several differences. Firstly, we have chosen to focus on a different empirical domain, oncology. All medical professionals attending our guideline group meetings were specialists, albeit from different disciplines. This probably accounts, in part, for the absence of a structuring effect of professional parameters on group dynamics. Oncology, moreover, has a long multidisciplinary tradition, which, in the French case we studied, is entrenched both in the institutional nature of comprehensive cancer centers and in state regulations. Secondly, the role of material and textual artifacts in the shaping of judgments and actions is a key element of pragmatic sociology, but this is notably absent in Moreira's analysis. In line with science & technology studies' longstanding focus on textual inscriptions and translations (e.g. Latour, 1990), and following up on Mykhalovskiy and Weir's (2004) programmatic suggestion to investigate the textual dimension of EBM, we pay special attention to textual practices. Guideline group meetings cannot be reduced to an exchange of arguments to select a winning position, after which the actual writing of the guideline would amount to a mere formality. Textual activities do not happen *after* consensus has been reached, they are *part* of the debate. Closure of debate does not necessarily imply that participants share the same opinion or interpretation. The collective production of a text — i.e. of a specific sequence of sentences and paragraphs that group members constantly readjust and rearrange until a final version is agreed upon — signals the end

of debate. Thirdly, we have borrowed from the sociology of organizations with respect to procedures and organizational routines. One of the most striking features of the dynamics of a guideline development group lies in the role of (local) procedures, rules and distinctions, as set by the guideline developing institution and flexibly enforced, interpreted, adapted and modified by group members. Our analysis will focus precisely on this "apparatus" (or *dispositif*, to use Foucault's notion) and, in particular, on the organizational and methodological routines that are deployed in the course of group activities.

Our focus on *dispositif* and texts has led us to an additional point. Clinical trials do not test substances nor do agencies such as the FDA approve them; both institutions test and process specific *claims* about substances. Amounting to carefully crafted textual statements about the scope and results of a clinical trial (e.g., substance X works against condition Y affecting patient population Z), claims are excerpted from publications, submitted for drug market approval and embedded in guidelines. As we will see, this is far from a mechanical transposition, but this process presupposes and depends on the upstream production of specifically formatted textual claims. As a result, guidelines no longer appear as self-contained evidence-based tools targeting individual clinician's behavior; they are elements of a chain of textual *translations* linking knowledge production about therapeutic substances and pathological processes, drug marketing and the regulation of medical practices. In other words, they partake in the emergence and stabilization of a new biomedical configuration grounded in the establishment of mutually constitutive links between two processes: on the one hand, the re-formatting of clinical trials into a device for producing carefully monitored evidence statements targeting specific markets (Greene, 2007) and, on the other hand, the increasingly pervasive role of regulatory processes within biomedicine (Cambrosio, Keating, Schlich & Weisz, 2006).

Material and Methods

Our ethnographic analysis centers on an oncology guideline development group convened by a French program called “Standards, Options, Recommandations” (SOR). Established in 1993 by the National Federation of the French comprehensive cancer centers (FNCLCC), with additional financial support from a national charity, the French National League against Cancer, and the government’s Health General Directorate (HAS), the SOR program was given the mandate to develop and update oncology guidelines in order to harmonize “clinical practices between cancer centers concerning diagnostic, classification, treatment and follow-up procedures” (Fédération Nationale des Centres de Lutte Contre le Cancer, 1994: 50; our translation). The FNCLCC is the umbrella organization of the 20 regional comprehensive cancer centers, whose origin goes back to the 1930s (Pinell, 2002) and which combine clinical research and treatments within a multidisciplinary framework. In 2008, the French National Cancer Institute (INCa, established in 2004) took legal responsibility for the SOR program. The program relies on a distinctive framework for the production of guidelines that emphasizes the need to follow the tenets of Evidence-Based Medicine by producing recommendations resting on the best available scientific evidence or on expert consensus when adequate evidence appears to be lacking (Fervers, Hardy & Philip, 2001). Between 1993 and 2006, SOR published 81 guidelines, i.e. 54% of the 148 published French clinical practice guidelines (Castel, 2009). Professional and public bodies have formally endorsed the SOR programme and medical audits conducted by Social Insurance use them as reference. The SOR guidelines have been diffused outside France, namely in the British Journal of Cancer, and the SOR program was one of the founding members of international initiatives such as the Guideline International Network.

Our research strategy was to follow the development of a particular guideline from the initial stages to the final drafting and circulation of the guideline document, a process that in our case took place in 2007-2008 over a period of 21 months. We selected a therapeutic guideline centered on a trans-disciplinary medical condition that affects patients with different forms of cancer. The study was approved by McGill University's Research Ethics Board I. For confidentiality reasons, we have removed any information that could allow readers to identify the specific topic of the guideline and thus individual group members. The group included clinicians from the two core medical specialties treating that particular condition, as well as medical oncologists, pathology/laboratory specialists, academic researchers investigating the condition, anesthesiologists, and two SOR methodologists, for a total of 22 participants. LK and HC attended and recorded guideline development meetings that were transcribed in full and jointly analyzed by all the co-authors. In addition to field notes, we collected the documents used in the development of the guideline, including the original and modified versions of PowerPoint slides, successive drafts of the guideline, comments of external reviewers and the group's responses to those comments. We interviewed the methodologist who was responsible for selecting and reviewing the literature, drafting the recommendations, and moderating the guideline development group meetings. (Since all our original material is in French, all translations are ours.) Last but not least, we could count on a considerable amount of background information and "experiential data" given that the three French co-authors of this paper have many years of acquaintance with the SOR program: PC has written a doctoral thesis on French cancer guidelines (Castel, 2002), BF has acted as director of SOR for several years and HC was employed as a sociologist by SOR.

Clinical oncology is, in a sense, pre-formatted for guidelines, given its reliance on protocols derived from large, multicenter clinical trials — a large proportion of which are carried out by non-commercial cooperative groups

—, and the aforementioned multidisciplinarity. The SOR procedures analyzed in this paper, however, are by no means unique to this program nor are they specific to oncology. We therefore expect our analysis to apply to other guideline development programs, in particular those sponsored by public or professional organizations.

The SOR *dispositif*

During guideline development meetings, the methodologist (see below) and the coordinator (or chair) constantly reminded group members of the procedural and methodological routines guiding and structuring the guideline development process. As mentioned in our introduction, observers have overlooked the extent to which social interactions and arguments are embedded in procedural rules and methodological devices that act simultaneously as flexible constraints and as resources. We can borrow (and slightly distort) Foucault's (1994: 299) notion of a *dispositif* — defined as a system of relations that can be established between heterogeneous elements such as discourses, institutions, regulatory decisions, laws, administrative measures and scientific statements — to portray SOR as an apparatus that associates and aligns participants, procedures, texts, topics and resources. In contrast to procedural rationality analyses that tend to grant a determining role to procedures and blur the contribution of participants into the background, we conceive of both methodological procedures and participants as constitutive elements of the *dispositif*; one does not take precedence over the other.

Participants include professionals such as methodologists and research librarians who are part of the SOR staff, as well as outside clinicians and experts from the relevant domains. Methodologists derive statements from the published articles and use technologies such as PowerPoint slides and Excel tables to present participants with summaries

and overviews of the existing literature. These tools provide an interactional infrastructure: debates take place around the projected slides, on which textual amendments are entered in real time. Two central notions, discussed below, are used to discuss, modify and rearrange statements: *Levels of Evidence* and *Standard/Options*. They allow group members to fine tune their recommendations so as to translate the degree of (un)certainty and (dis)agreement that surrounds certain procedures. A number of explicit and implicit procedural rules and categories are used to operationalize these two notions.

Participating physicians come not only from the network of cancer centers to which SOR was affiliated, but also from public and private hospitals. By the turn of the century, about 2,000 practitioners had been involved with the SOR guidelines, half of them from the Federation's cancer centers. SOR can thus claim to staff its guideline development groups with the best experts, regardless of institutional membership. Methodologists, in contrast, are project managers hired by SOR to prepare a synthesis and critical appraisal of the evidence and to organize the guideline group meetings. Originally trained as biostatisticians, information specialists or health professionals, methodologists are usually not MDs; some of them are, but they no longer practice and are thus unable to claim clinical expertise. Still, they are full-fledged participants in group discussions, officially entrusted with the task of ensuring that the resulting guidelines remain congruent with the available evidence. In 2000, because of the growing amount of clinical publications and the complexity of managing the resulting databases, SOR recruited research librarians to support the methodologists' work.

Systematic reviews of the scientific literature often show that evidence on a given procedure is either absent, of limited quality, contradictory or derived from a different patient population, clinical setting or procedure. The SOR *dispositif* thus also elicits and encompasses expert

opinions. As the methodologist explained during an interview, “when many data are available, it’s a no-brainer, data speak by themselves … when we have to deal with less data, it’s up to the expert to speak up”. To counter the potential liability caused by this situation, the formal SOR methodology suggests that the members of the development group ought to be chosen “by the most objective criteria” (Dosquet, Goldberg & Matillon, 1995: 759). “Objectivity”, in this context, is defined as a “civic worth” (Boltanski & Thévenot, 1999) insofar as it refers to the inclusion of members from a wide variety of specialties and groups, public as well as private institutions, at least one rank-and-file clinician and sometimes nurses and patient representatives (in our case the latter two groups were absent). While professional organizations can and, in our case, did formally endorse the procedure by jointly supporting the guideline development project and by approving the resulting guideline, they did not elect or select experts from their midst to act as their official representatives on the guideline committee. Clinical coordinators recruited them from amongst their professional and personal contacts on the basis of their expertise and national visibility. As a methodologist explained in an interview, “they are colleagues that they meet at meetings and workshops, with whom they have exchanged or co-authored articles. … One takes the two best French [name of specialty] and the two best [name of specialty], and that’s it”.

As can be easily seen, the SOR *dispositif* follows the tenets of EBM by relying on literature reviews and published evidence, but has also adopted a pragmatic approach by supplementing formal methods with a (necessary) reliance on expert opinion for the assessment of that information. The official endorsement by professional organizations, and the presence of experts from the relevant specialties within the development group, increases the legitimacy of the procedure. It also facilitates the practical requirement of promoting the guideline within each of the experts’ distinct professional and geographical milieus (their “parishes”). SOR

expects members of a guideline group to act as the guideline's advocates and promoters: "we have here representatives of specialties from very different horizons, we have to think how to communicate [the recommendations] to the colleagues in our own parishes because, first of all, there is a time issue, the issue of being recognized [as experts in the field] and, more importantly, there is the issue of the legitimacy of this group"[4].

External review of draft recommendations is another important part of the SOR *dispositif*. They are generally mailed to 150-250 external reviewers (230 in the present case) from different specialties and domains. Presented as a form of external quality control testing, this "national review" process acts as a way of internalizing the external world and thus of fostering the acceptance and implementation of the final recommendations by potential users. In practice, only 25-30% of the reviewers (28% in our case) actually provide comments and suggest modifications. These feedbacks often lead to changes of the actual content of the guideline. Equally as important, and in spite of the relative small percentage of respondents, the review is used to demonstrate the approval of the medical community at large. The final guideline document includes tables and pie charts stating, in our case, that 93% of consulted physicians approved the guideline and 94% said they would use it.

The *dispositif* is what makes the process accountable, in the dual sense of providing participants with elements and rationales for defining and modifying the guideline's content, and for establishing the legitimacy of those activities. Not only should procedures be followed, but this should be done in a visible and transparent way: search keywords, evidence tables, external review results, the names and specialties of participants are all listed in the guideline or its appendices. The *dispositif* itself gains its legitimacy from this transparency, its institutional entrenchment in the

---

[4] When not otherwise indicated, all quotations are from meeting transcripts.

FNCLCC and its interfaces with international initiatives. It is too early to know whether and how SOR's takeover by a government institution (INCa) will change the situation.

Procedural rules and categorizations

At the beginning of every guideline development meeting, the methodologist shows two PowerPoint slides to remind group members of the categories and procedures that are supposed to guide the process. By accepting to become members of the group, participants are expected to abide by these "rules of the game". Confronted with a participant who broke this implicit contract by questioning the basic procedures, a methodologist reacted rather swiftly by stating that "the approach we take is the classical SOR approach, to which we all subscribed … the fact of being a group member, and we thank all of you for that, means that you accept to play by the rules of the game, we all accepted it at the outset, and it's a clear methodological rule".

As we will see below, acceptance *in principle* of these rules does not mean that they cannot be challenged *in specific instances*. It is useful to refer, in this respect, to the distinction introduced by Feldman and Pentland between ostensive and performative aspects of routines. Borrowing a classical Wittgenstein argument, they argue that "no amount of rules is sufficient to specify a pattern of behavior fully, because the interpretation of any rule, or any part of a rule, requires more rules" (2003: 101). Much of the group meetings we observed centered on how to apply rules to particular cases, i.e. how to perform ostensive routines. For instance, while participants had to agree that the literature review that provided a basis for the guideline rested on a formal assessment of the type of evidence reported by articles, they were also able to point to details of specific studies that cast doubts on their conclusions (or even mention extra-textual

contingencies: “I’m fed up with [study X] … they keep trying to fob this off on us day after day, this has been going on for twenty years, I’m fed up”). They were thus able to reject or reduce the evidential status of claims that, formally speaking, ranked as high-quality evidence.

Our case study reinforces Feldman and Pentland’s (2003) argument, showing that the procedural rules provided by SOR amounted to a mixture of *both* constraints *and* resources (see also Crozier, 1964): constraints because “ostensive” rules and routines had to be taken into account in group decision-making, and resources because they were flexibly “performed” to generate decisions. In other words, the SOR procedures were not a mere varnish that was used retrospectively to legitimate decisions taken on other grounds, they were mobilized in the course of discussions and made a difference. Neither did they structure these discussions in any deterministic way or lead to pre-established conclusions.

So, what were the rules and procedures displayed on the two introductory slides? The first slide introduced the distinction between Standards and Options, obviously two central categories insofar as they are embedded in the institution’s name. The second slide defined the four main categories of Levels of Evidence that participants had to use in assessing the evidence and drafting recommendations.

Levels of Evidence

The “Levels of Evidence” (henceforth LOE) slide listed four levels (A to D) ranging from evidence supported by “good quality meta-analyses” or by a “coherent” set of “good quality” randomized clinical trials (level A), to no data or only case report evidence (level D). Levels contained sub-categories: the B level, for instance, included B1 evidence (randomized trials) and B2 evidence (prospective or retrospective studies). Resort to an evidential hierarchy is of course not a SOR innovation; it is a defining feature of EBM and a core methodological component of systematic literature reviews.

There are, however, dozens of slightly different hierarchies, in spite of the fact that an international collaboration has attempted for several years to devise a single standardized version (GRADE working group, 2004).

As previously mentioned, one of the participants launched a wholesale attack against the hierarchy used by SOR and, one could argue, against the very idea of establishing formal rules for determining LOE, since his target was the failure to provide a detailed assessment of each study: "the methodological quality, the type of randomization, the non-existence of treatment differences between the two arms of the trial, the analysis that led to the reported results, the criteria that have been taken into account, and so on … all these elements are totally absent [from the LOE scheme]". The methodologist and the clinical coordinator, both acting as spokespersons for the SOR *dispositif*, rejected this global challenge on pragmatic grounds: the need for consistency across SOR guideline groups ("this particular hierarchy has been used for all the guidelines in the past 10 years", "we cannot change LOE tables every six months for each project") and the fact that in the absence of international standards the LOE used by SOR were as good as any other:

> *Coordinator:* This being said, all these [LOE classification] tables can be criticized: I had a look at the tables used by the Italian guidelines, and then I looked at the tables used by the [name of French healthcare agency], and I said to myself, gosh, there is no coherence, and then one has to decide [which one to use], and so we might as well use the SOR tables…

When the challenger insisted with his criticism, the methodologist became more blunt: "those are the rules of the game, we don't modify them … that's how it is, you can't do anything about that". This did not mean, however, that the attribution of a LOE category would proceed automatically: to the contrary, once the general principle was accepted, its application to

individual studies often raised many discussions. Did a given study amount to a “good quality” clinical trial? Did it count as a prospective or a retrospective study? If four out of five studies agreed, did that count as “a coherent set of studies”? In the case of particular studies, participants discussed criteria such as the number of patients included (“Either we stick with B2 or if we consider that too few patients were included, we stay with C”) or the date of publication (“So, let’s use only recent studies, because if we exclude [the older reference X], then the conclusion can be level of evidence A or B”), or even distinctively more idiosyncratic parameters such as the previously quoted statement about being “fed up with [study X]”.

Quite often, discussions focused on how to categorize a given study’s methodological design. For instance, a study claimed to be a retrospective analysis, but when one of the participants asked what that really meant, a long discussion followed which did not lead to consensus:

> *Methodologist:* Well, that means that they took the [study] arms a posteriori, the five studies …, they extracted data from the five retrospective studies.
>
> …
>
> *Participant 2:* I think that we should say: “it’s an a posteriori analysis of five prospective randomized trials” and add that it’s an a posteriori analysis, with cancer patients, of five studies. And in fact this is what it is, we are mixing up prospective and retrospective.

After much discussion, and in the interest of time, the coordinator put an end to the debate by deferring the decision to the methodologist: “OK my dear friends, we have to go ahead… I’ll write down in red ‘verify methodology’ and we will check that”.

As it can be seen from these selected but representative examples, procedural rules and distinctions provided a frame for the debate, but they were unspecific and flexible enough to require interpretation in their

application. They were in fact the starting point and a recurring element of the debates that took place within the group and were embedded in the justifications offered in support of a given statement, but they did not dictate the outcome of interactions. The methodologist and the coordinator played an important role in reminding participants of the official frame but the debate was essentially open-ended. When participants contradicted the assessment by a methodologist of the quality or LOE of a given study, they did so *thanks to*, rather than *in spite of* the existence of procedural rules. The application of the rules to specific cases was open to debate and enforcement was limited to frequent but friendly warnings that could be quite elastic: "I remind you that [in this recommendation] there is no level of evidence and, it seems, we step outside our role and our job, *but oh well!*" (Methodologist, our emphasis). Boundaries were elastic, and so, by redefining them, one could step outside them without officially doing so. But it would be wrong to infer that "anything goes", for these actions had to be justified: there were rules, routines and categories (however elastic) that *had* to be heeded. The specific justification that would eventually carry the day, however, emerged from the debate and was in no way pre-determined.

Standards and options

In addition to LOE, a second important device contributed to the open-ended framing of the debates, namely the distinction between Standard and Options. According to the first introductory slide, a Standard refers to a clinical procedure that experts *unanimously* regard as the gold standard. Options refer to clinical procedures (notice the plural) that experts consider *appropriate*, even in cases when experts favor one of those options. It is important to note that, unlike LOE, the Standard/Options distinction is based on the *level of agreement*, and not on the *level of evidence*: standards are by unanimous decision, while options are agreed upon by a majority (Fervers, Hardy & Philip, 2001). *Ideally*, a standard should be based on the

highest LOE, but this is not a necessary condition: as noted by a methodologist in response to a commentary from an external reviewer: “A recommendation can be listed as a Standard even in the absence of data from the literature or even when supported only by a weak level of evidence”.

This flexibility to recast a statement as either a standard or an option was a very useful device for bringing (temporary) closure to a debate by allowing recommendations to be made even in the presence of disagreements and uncertainty. After all, as mentioned by participants, it was precisely in the areas where evidence and unanimity was lacking that they were expected to provide guidance for practitioners: “those who read our recommendations, they expect from us that we give them options, even weak options based only on expert opinion, and if we don’t give them options, it’s not worth doing all this work”.

The Standard/Options distinction was a tool for managing uncertainty in an additional, more subtle sense. It helped to define the “grey zones” in need of further investigation by redefining the boundaries of uncertainty as new arguments were introduced about the availability or absence of evidence, and by fuelling discussions about the tasks that lie ahead: how to learn what was not yet known, where to look for possible answers. Uncertain domains were not zero-knowledge domains or independent variables constraining medical activities or judgments, as Renée Fox (2003) would have it. Rather, uncertainties were the outcome of the actors’ practices (Bourret & Rabeharisoa, 2008).

Some participants actually understood the distinction between Standard and Options in terms of *degrees of certainty* (and not in terms of *agreement*, as previously mentioned): standards were certain, options had not been proven beyond reasonable doubt. The trouble with this interpretation was that it left open the possibility of having several coexisting standards (with well-established criteria for choosing between them) or of

ending up with a single (uncertain) option, contradicting what others considered the defining plurality criterion. Witness the following exchange:

> *Participant 1*: But that is a standard. It's a standard to say that we can't say anything; it's clearly not an option.
>
> *Participant 2:* No, a standard and an option, they should both recommend a given procedure, so it really should say "I can or cannot use [drug X]".
>
> *Participant 3:* If it's optional, by definition we need at least two possible procedures ... there needs to be several alternatives.

Recommending several courses of action raised the question of how to define the criteria for deciding between the available choices. One way of handling this question was to leave the choice to the individual clinician facing a given clinical situation (use option X "if needed"), but other participants insisted that the recommendation should specify the conditions under which a given option became quasi mandatory ("We must list situations in which an option should be used"), thus turning it into a proto-standard.

Both in the case of LOE and of the Standard/Options categories, it is tempting to resort to the category of *negotiations*, as sociologists often do, to account for the interactions we reported. A number of situations we witnessed appeared to fit such a description, for instance in the case of long exchanges about whether the LOE of a given claim should be moved up from B2 to B1 or, to the contrary, retrograded to C. And yet, negotiations is not the right term, since it implies that the issue was well defined, that the negotiating groups had preexisting opinions or interests about it, and that they looked for a more or less favorable compromise, possibly in exchange of other concessions. The situation in the guideline group was different: in the case of controversial statements, it did not amount to a clash between pre-established choices or opinions (underwritten by different professional

interests) but to an attempt to chart uncharted territory. As discussed in the next section, the process did not imply a freewheeling discussion: framed by the SOR *dispositif*, it focused on the fine-tuning of *textual statements* to make them compliant with two requirements: that they form a coherent sequence of interconnected textual statements, and that they include the appropriate mixture of precision and vagueness.

Textual practices

So far, we have discussed the constitutive elements (people, procedures, rules and categories) of the SOR *dispositif* and the dynamics they engendered. It is now time to focus on the *product* of these activities. Offering a forum for the exchange of arguments, or reaching an intellectual consensus on the value of a given intervention was not the final purpose of this complex procedure. Rather, it was to produce a text. The chain of textual translation began with the extraction of textual claims from published clinical trial reports: methodologists compiled extracts and created tables and summaries. Several of these statements were translated without much debate from evidence tables to conclusions and turned into recommendations, sometimes literally by copy and paste. In a significant number of cases, however, textual statements led to more or less heated debates within the group. Without discounting the presence of the former occurrences, we focus on the latter.

By following closely the debates that took place within the group, we noticed that it was sometimes unclear which argument had won the day, or how divergent arguments added up to a given conclusion; yet, in the absence of objections to a proposed text, closure would still happen. To produce a text members of the guideline group did not need to agree on a given matter. A textual statement could be formulated in such a way as to leave room for flexible interpretations; group members would then feel

comfortable with underwriting it even if its exact operational meaning remained unclear. For example, some participants argued that a given treatment should be administered for three months, while others opted for six months. Each group maintained its position and no intermediary solution (four or five months) seemed viable. Closure was reached around the following formulation: “at least three months”. In this way, both options were consistent with the guideline. Having reached consensus on that formulation, group members were able to walk away with divergent opinions.

Several observers have claimed that the textual clarity and precision of a guideline have positive effects on its implementation (e.g. Michie & Lester, 2005). But, as already mentioned, lack of precision is often *needed* to produce a text providing guidance, for controlled fuzziness is a resource for generating closure on a statement (Allen, 2009). There are also other reasons why group members considered it impossible or inappropriate to provide detailed guideline statements, such as a practical impossibility to foresee the degree of detail needed by every practitioner, patient and institution operating under different circumstances. Specifications could be made later, at the local level during the implementation phase. A boundary was drawn (in principle) between the job of developing a guideline and the job of implementing it:

> *Coordinator:* … we have a [guideline] text, and its application belongs to a subsequent stage, and for that there will be [the equivalent of] pull-out recipe cards or summary cards that will become available and people will produce them in a variety of forms. Because we cannot be expected to write down bedside prescriptions. Therefore there will be, we will all be asked to be part of a six month period during which we will have to explain the [guideline] text.

The formulation of the text of the SOR guideline was a truly collective effort. Methodologists or individual group members did not write the text

after the meeting according to their recollection and understanding of the debates, instead, the *precise* wording of the text was debated and worked out in real time on the PowerPoint slides. The issue was not one of editorial esthetics; rather, the choice of words was held to define the content and the (possible) consequences of the guideline recommendations. Moreover, to produce a guideline the group had to decide not only *what* to write but also *where* to locate specific statements in the document. The text of SOR guidelines follows a strict, pre-determined sequence that suggests a linear process according to which the literature provides evidence associated with LOE, which is assessed and/or supplemented by the experts, all this leading to a set of recommendations. The sequence did in part correspond to the actual processes we observed (e.g., the methodologist prepared a literature review before the first group meeting), but the group discussions were far from linear. The positioning of specific statements within the overall organization of the document did not automatically flow from the layout conventions. During the discussions, group members travelled back and forth between sections:

> *Participant:* … who says that? The literature says that? Or it's you, based on your clinical practice? … If it's you, it should go in the Expert Judgment section.

Travelling between sections also implied revisiting phases of the guideline development process. For instance, participants confronted with a particularly thorny issue agreed on the need to draft a strongly worded recommendation to prevent widespread instances of "bad practices". But a "strong" recommendation should ideally be based on "strong evidence" which, in the present case, was not available when participants reviewed the relevant publications. So, they revisited them and, by eliminating reference to a particular study whose results were not in agreement with the other studies and whose methodology seemed doubtful to some participants, were

able to raise the overall LOE for the recommendation from C to B1. Although the literature review preceded the drafting of the recommendations, in this particular instance it was retrospectively modified to make it consistent with what participants felt was a crucial recommendation.

SOR guidelines must include both a set of specific recommendations and the reasons for those choices. The textual location of a statement played a critical role in establishing evidential connections between the recommendations and the different kinds of justifications supporting them. Group members were reminded of the fact that the coherence of a guideline depended on its overall organization and the need to showcase this coherence:

> *Methodologist:* The reader must get a clear sense of all this, he should not be left wondering how we ended up with this recommendation, like "there was no literature and here they go and suggest a standard", see?

Skeptical or critically minded readers might infer, at this point, that the final layout of the guideline, with its neat argumentative sequence, was "merely" a rhetorical device for retrospectively justifying decisions and recommendations that were taken on different grounds and certainly not by following the linear process implied by the layout. Yet, this conclusion would miss the decisive fact that the linear template acted as a performative device prompting group members to move back and forth between sections of the text and to adjust its coherence little by little.

Coming to terms with the external world

The avowed objective of guidelines is to change healthcare practices by replacing substandard interventions. While maintaining a proactive stance, guideline developers are also aware of the fact that they need to adjust

recommendations to prevailing external conditions. How and to what extent remains an open question, and leads to discussions about the potential impact (or lack thereof) of specific recommendations. By evoking possible scenarios, guideline developers attempt to manage relations between recommendations, users and other relevant stakeholders.

In this respect, the aforementioned distinction between production and implementation proved to be a useful tool for framing the participant's work, but was often difficult to maintain and had to be re-specified when group members were confronted with the practical task of wording specific recommendations in such a way that they could have a reasonable chance of being of any practical import. Participants, arguing that they were "not going to recommend something that has no chance of being applied", wondered about what would make health practitioners "out there" more likely to modify established routines and replace them with the new recommendations. The obstacles they mentioned included patient preferences ("patients don't want that"), a lack of local resources ("the necessary radiological equipment must indeed be available") or legal arrangements ("If we write that down in our recommendations, there will be only 20 centers in France that will have the right to [perform the procedure]"). While this kind of concerns were frequently raised, they were often declared off mandate or unacceptable, at least in principle:

> *Participant:* If we argue that [a procedure] should not be performed because there is no evidence, OK, I agree with that. But to argue that "this will not be accepted, this will not gather any support, it will not be done because patients don't want that" … I cannot agree with this kind of argument.

Group members, insofar as they were health practitioners, had a dual status: they were experts but also, and simultaneously, users, albeit "expert users" as contrasted with "rank-and-file users". They were thus able to "switch hats"

during debates. Role switching did not correspond to specific sequences of the guideline production process (e.g. experts during LOE discussions and users when debating practical issues), but took place at any stage in the process. Participants, however, considered the role of state regulators or administrative decision-makers beyond their purview. To steer clear of conflict, guideline recommendations should as far as possible avoid addressing the concerns or mandates of other jurisdictions:

> *Coordinator:* It's up to them [French drug regulators] to get a move on and make [a particular version of a drug] available in France, should they judge that it is appropriate. It's not our problem.

In the following discussion, group members, while agreeing on which drugs should be used, were faced with the fact that a particular drug has not yet won regulatory approval in France. They thus wondered whether they should insert a statement mentioning its "limited availability". This issue led to much soul searching: Would the recommendation have a chance, under such circumstances, to be correctly implemented? Could this be interpreted as favoring a drug over another or as meddling with approval procedures, and, if so, how would the pharmaceutical industry react? Would this affect the perception of the integrity and transparency of the guideline development process? Group members constantly reminded themselves that they would be held accountable for their work and thus of the need for a delicate balancing act so as not to antagonize a range of different parties.

The guideline development process also included a formal mechanism — the aforementioned external review — to align the guideline's content with the "external world". This procedure is part and parcel of the *dispositif*, in the dual sense of being one of its formal requirements and of supplying feedbacks, both real and apprehended, leading to text modifications. Awareness of the impending review clearly shaped the wording (and thus the content) of the recommendations. The following

remark by a methodologist was typical: “[These data] should allow us to opt without major contradiction for a B2 level of evidence, and without risking a full-fledged attack when [the draft guideline] will be circulated for review”.

A guideline is a text but, unlike, say, the text of a scientific article that, once published, severs its ties with the laboratory from which it originated, the text of a guideline remains attached to its producers after its publication. It will be revised at more or less regular intervals to take into account new evidence, publication of the final text is followed by a period during which guideline developers promote the results of their work to potential users, and, last but not least, recommendations will subsequently have to be rendered into summaries, decision trees and regional guidelines adapting them to local conditions, thus pursuing the cycle of textual translations.

Conclusion

Increased international collaboration and comparisons have led to the somewhat disheartening observation that guidelines on the same topic frequently contain divergent recommendations. Many guideline developers and EBM advocates argue that this is not necessarily a problem insofar as variation is warranted by different local (usually national) differences and certain procedural requirements are fulfilled (Eisenberg, 2002; Fervers et al., 2006). The most widely used instrument to evaluate guideline quality does not examine a guidelines’ clinical content or its underlying evidence, but only evaluates the characteristics of its production process (AGREE Collaboration, 2003). In short, production procedures are increasingly important to legitimate guidelines. In contrast to formal, checklist-like evaluation tools of the guideline developmental process, our paper focused on the ethnographic observation of how a particular guideline *dispositif* framed the production of a guideline, calling attention, in particular, to the textual dimension of these activities. While subject to limitations due to its

restricted focus, this kind of detailed analysis has the potential to contribute to a better understanding of the dynamics of guideline production, especially if combined with similar studies carried out in other settings (e.g. Moreira, 2005). We hope, in particular, to promote studies of the *dispositif* of other national and international guideline development institutions, especially those mandating public or patient participation, involving commercial in addition to professional actors, or outsourcing evidence-gathering processes.

Our analysis highlights that a guideline's credibility does not rest solely on its evidential basis, as quality assessment of evidence is an intrinsic part of the process of developing guidelines. Moreover, this process is in part, self-vindicating, insofar as it is made possible by the existence of meta-regulatory documents such as other guidelines. But nor is its objectivity based on mechanical procedures for compiling facts and data. Its legitimacy rests on the articulation of heterogeneous types of expert knowledge and judgments, both within the guideline development group, and vis-à-vis an external world of textual documents (clinical trial reports, other regulatory texts, external review comments, etc.). As such, guidelines act as mediators at an important junction in the extended chain of textual statements produced, assessed and processed by a range of public and commercial research institutions and regulatory agencies that are part of the meta-regulatory web of contemporary biomedicine (Cambrosio et al., 2006). As mentioned in the introduction, clinical trials, the basis on which the EBM enterprise rests, are not so much devices for testing drugs in any "absolute" sense, but, rather, have become devices for producing specific claims concerning a given molecule, test or procedure (Greene, 2007). Once excerpted from the original publications, these formatted statements can be further processed (e.g., by associating them with a LOE) and transferred to other texts, such as guidelines. This, however, is not a mechanistic account,

insofar as the role attributed to expert opinion (as embedded, for instance, in the Standard/Options distinction) acts as a versatile counter-mechanism.

It would be wrong to portray this counter-mechanism as a form of resistance (for instance, by clinicians) to the reconfiguration of biomedical practices, in the same way as it would be incorrect to equate the work of methodologists with the *dispositif* and pit it against the request for an "alignment with the external world" conveyed by clinicians. "Expert opinion" and "evidence-based statements" are part and parcel of the same *dispositif* and work together to produce new knowledge and new practices that do not correspond to the somewhat paranoid picture of a corporate hijacking of medical knowledge but neither do they correspond to the positivist utopia of an EBM free from any undue interference from anecdotal, contingent or even idiosyncratic forms of justification.

CHAPTER 4

Evidence Searched Guidelines: How to manage an absence of evidence in Evidence Based Medicine

Assessing the absence of evidence

On February 28th 2007, the Program in Evidence-Based Care (PEBC) held its first scientific symposium to discuss the "Nature of Evidence". PEBC develops guidelines for Cancer Care Ontario and is affiliated with the department of Clinical Epidemiology and Biostatistics at McMaster University, the 'birthplace' of Evidence Based Medicine (EBM). The biggest future dilemma facing EBM guideline developers identified at PEBC's symposium was how to manage the *absence* of Evidence: "how to address the challenges of providing *evidence-based* advice to address questions for which the *evidence is lacking*, of poor quality, immature or incomplete" (Program in Evidence-Based Care, 2007: 5, emphasis added). This article takes as its starting point this seemingly impossible challenge of how to 'be evidence-based' in the absence of evidence.

The absence of evidence provides a productive starting point to contribute to social science literature on the nature of evidence in medicine. The incompleteness of knowledge is not simply a void about which nothing can be said, as Knorr-Cetina's notion of 'negative knowledge' already showed. Scientists gain valuable information about an object by finding out what it is *not*, and by identifying errors and uncertainties (Knorr-Cetina, 1999: 64). Within medicine, scholars have detailed how 'failed' medical experiments can produce successes (Timmermans, 2011), and how pharmaceutical companies can benefit from uncertainty (McGoey, 2009) and

how the establishment of uncertainty can be considered a useful outcome of research, standardization and regulatory activities (Cambrosio et al., 2009: 659; Rabeharisoa & Bourret, 2009; Moreira, May & Bond, 2009). Far from singular and nondescript, "nonknowledge is constructed, assessed, and communicated in contrasting or even incompatible ways", an absence of knowledge can be conceptualized as a variant of knowledge (a known unknown, calculable risk, or not-yet-known), the entirely unknowable (unknown unknowns), or ignorance (intentionally unknown) (Böschen et al., 2010: 803). Examining how the absence of evidence is defined, assessed and managed within the EBM paradigm can teach us about the nature of evidence as well as about the nature of EBM itself, as EBM's definition of evidence is considered its defining feature.

Evidence Based Medicine prominently claims to be a 'paradigm' shift in medicine, away from "intuition, unsystematic clinical experience, and pathophysiologic rationale" towards randomized clinical trials (RCTs) as a more scientific basis for medicine (Evidence-Based Medicine Working Group, 1992: 2420). By providing clinicians with 'evidence', delivered to them by systematic reviews and guidelines, EBM aims to transform and rationalize the way uncertainty is controlled in the clinic (Armstrong, 2007). This transformation has been characterized as (and criticized for) a replacement of the 'subjective' - but personalized and contextualized - expertise of individual clinicians with the 'mechanical objectivity' (Daston & Galison, 2007) of automated rules and universal evidence of a standardized medical science (Cronje & Fullan, 2003). Empirical observations of the use of EBM guidelines to handle uncertainty in the clinic do not find the knowledge and skills of clinicians is simply replaced by the 'order' of science and standards (Timmermans & Berg, 2003). Clinicians continue to rely on diverse, incomplete and local knowledge from a variety of sources (Armstrong, 2002; Latimer et al., 2006; Rabeharisoa & Bourret, 2009). This is not simply a matter of clinicians 'resisting' or 'failing at' EBM, nor of

clinicians favoring a competing paradigm such as ‘patient centered care’ (Armstrong, 2002), but is what EBM looks like in clinical practice. Even its founding fathers redefined EBM as “the integration of the best research evidence with clinical expertise and patient values", acknowledging diverse knowledge is needed when *applying* evidence to individuals in the clinic (Sackett et al., 2000: 1). With this broadened definition of EBM the practice of EBM no longer seems such a ‘paradigmatic’ change, yet EBM’s *definition* of evidence remains novel and specific. EBM’s now infamous Evidence Hierarchy ranks evidence according to the methodological design by which it was produced: from case studies (the lowest level), to non-controlled cohort studies, to meta-reviews of randomized and double blind clinical trials (the highest level). This hierarchic classification is based on a study’s internal validity (whether its findings are likely replicable), rather than a study’s external validity (whether a study’s findings are valid and useful in settings outside the trial), also referred to as applicability or relevance (Grossman & MacKenzie, 2005; Cartwright, 2007). Here I will refer to the EBM-specific characterization of evidence as Evidence (with capital E) to distinguish it from the common notion of evidence that includes more diverse kinds of knowledge, proofs and justifications. Notwithstanding EBM’s support for a diversity of knowledge in the *application* of Evidence, the formal definition of evidence in EBM continues to prioritize quantified data from large patient populations produced by clinical epidemiology (Buetow & Kenealy, 2000). EBM’s classification of evidence has been strongly criticized for undervaluing many kinds of knowledge, such as alternative methodological designs (e.g. observational studies), knowledge produced by entire disciplines (e.g. public health, medical anthropology), as well as a disregard for solutions that are not easily tested in randomized clinical trials (RCTs) (Lambert, 2006; Grossman & MacKenzie, 2005; Black, 1998; Bluhm, 2005), and see the “fixed hierarchy of evidence as the guilty source of [EBM’s] questionable epistemic practices” (Goldenberg, 2009: 171).

By investigating how Evidence and the lack thereof is defined and handled in the development of EBM tools, this article contributes to the critical question "what counts as Evidence" according to EBM. It differs from much existing analyses of the nature of Evidence, which primarily consist of abstract and normative critiques of the formal Evidence Hierarchy. In contrast, I analyze how medical literature gets classified as Evidence in the practice of EBM guideline development. While clinical practice guidelines have multiple origins and can serve diverse purposes (Weisz et al., 2007), Evidence Based Guidelines have become such an important part of EBM that Timmermans and Berg effectively equate EBM with the use of guidelines (2003; 3). This study draws upon and further specifies some of the previous work on the pragmatic evidence work, new institutional procedures, types of justifications and participants mobilized in the production of a guideline text (Moreira, 2005; Moreira, 2007; Will, 2009; Knaapen et al., 2010). It specifically builds upon the previous chapter on the development of a French cancer guideline, but instead of addressing how distinctions *within* the Evidence Hierarchy are agreed upon, this article specifically focuses on the procedures that define Evidence as a category (regardless of its level); how Evidence is distinguished from other kinds of knowledge; and procedure to manage 'non-Evidence'.

After describing the empirical material this article is based on, the article consists of three main sections. The first introduces the term Evidence Searched Guidelines which better captures that the "EBM" label is not earned by the presence or absence of Evidence, but by the elimination of intentional ignorance. In the second section I argue that what counts as Evidence is not simply determined by the formal Evidence Hierarchy, but by the members of each Guideline Development Group (GDG) that agree upon specific criteria and put forward a range of knowledge, principles and categories to determine what counts as (relevant) Evidence. The third section describes the various kinds of 'non-Evidence' that are mobilized

during EBM guideline development, and distinguishes hierarchical and mediating relations between non-Evidence and Evidence. In the concluding sections I will discuss how the procedural rules of Evidence Searched Guidelines shift the notion of rationality promoted and produced by EBM. Since these procedures reveal, rather than expunge, non-Evidence, EBM guidelines are characterized by transparency, rather than Evidence. I argue that EBM does not produce 'mechanical objectivity' but reflects a reliance on 'regulatory objectivity' (Cambrosio et al., 2006), which is marked by the provision of procedural rules to handle evidence and consensus in the collective production of standards on how to proceed objectively. This suggests a specific inter-relation of the repertoires of evaluation mobilized in guideline development, positing the process repertoire as a meta-repertoire (Moreira, 2005). Finally, I will suggest the notion of transparency as a central issue to be further explored to understand the way the designers, producers and users of EBM tools regulate - and are regulated - within medicine.

Material and Methods

To understand how guideline developers understand and perform the nature of Evidence and EBM in their practices, this study principally relies on empirical material gathered from two guideline development organizations that promote and follow EBM principles: the Dutch Institute for Health Care Improvement (CBO) in the Netherlands and the Program in Evidence-Based Care (PEBC) in Ontario, Canada. The CBO collaborates with the Dutch Cochrane Center and The Dutch College of General Practitioners (NHG) since 1997 to establish methods for Evidence-Based Guideline Development (Evidence Based Richtlijn Ontwikkeling - EBRO), creating the EBRO platform in 2001. The EBRO guideline development manual was published by CBO, most recently updated in 2007 (EBRO, 2007). In Canada, the Program for Evidence Based Care has its origins at the

McMaster department of Clinical Epidemiology and Biostatistics, the birthplace of EBM, and in 1994 the "Practice Guideline Initiative" was integrated into what is now Cancer Care Ontario. A 'systematic' methodology called the "practice guidelines development cycle" was devised in 1995 (Browman et al., 1995). These guideline organizations produce each about a dozen guidelines per year that are highly regarded in their respective country and province, and both organizations train other guideline developers in evidence based guideline development through national and international workshops, meetings and summer institutes for novel guideline developers. The (former) directors and guideline developers of CBO and PEBC (Jako Burgers, Kitty Rosenbrand and Melissa Brouwers) publish abundantly about guideline methodology (including 9 articles co-published by JB and MB). They are founding members of the AGREE collaboration whose website[5] claims to "advance the science of practice guidelines" and that directly led to the establishment of the Guidelines International Network (GIN), of which CBO and PEBC are members. GIN aims to "improv[e] the efficiency and effectiveness of evidence-based guideline development, adaptation, dissemination and implementation" (GIN website[6]), and has developed international standards for guideline development (Qaseem et al., 2012). The practices of guideline development at CBO and PEBC examined in this study are institutionally distinct, but are comparable as their international collaborations inform and harmonize their Evidence Based Guideline principles, and also shape the methodology of guideline developers around the world.

The fieldwork on which this article is based was conducted over a period from 2007 to 2011. In addition to document analysis of medical literature, guideline development manuals and presentations that guide the

---

[5] www.agreetrust.org/ (accessed August 7th 2012).

[6] www.g-i-n.net/about-g-i-n (accessed July 15th 2011).

practices of Evidence Based guideline developers, I held twelve semi-structured interviews with guideline methodologists at CBO (6) and PEBC (6) that were recorded and transcribed in full. I participated in and observed training and networking events organized or attended by CBO and PEBC guideline developers, including five subsequent Guidelines International Network's annual meetings; two summer institutes (one on 'knowledge translation' and one on guideline development); one team meeting at PEBC; and worked as a guideline reviewer with AGREE. Detailed fieldnotes were taken at these events. I observed fourteen meetings of guideline development groups at the CBO in the Netherlands (on six different guidelines) and eight Teleconference meetings at the PEBC in Canada (on a single guideline). These were audio-recorded and specific passages of interest (based on fieldnotes) were transcribed. The notations used refer to the guideline group (GDG1, GDG2) and the number of meeting (a,b,c). I translated material that was in Dutch into English, and to maintain confidentiality of participants certain information in direct quotes taken from GDG meetings was altered (such as the guidelines' topic, the names of treatments or authors). The study was approved by McGill University Research Ethics Board I.

Being EBM without Evidence: Evidence Searched Guidelines

Clinical practice guidelines have been produced since the 1970s, by hundreds of organizations for a variety of purposes (Weisz et al., 2007). The founding fathers of EBM found most of those "non-expert 'expert' reviews" that were not "worth following" because they were based on the 'opinion' and 'consensus' of national experts (Sackett & Rosenberg, 1995: 632). As it became clear that individual clinicians lack the time and skills to critically assess medical literature themselves, EBM advocates did encourage groups

of experts to provide 'pre-appraised evidence', according to specific principles of evidence search, synthesis and assessment. Particularly the systematic literature reviews produced by the Cochrane collaboration, and guidelines based on evidence reviews (Guyatt et al., 2000).

Guideline developers themselves see a significant difference between their own products and the "evidence reports" that Cochrane produces, as guidelines are "recommendations for practice" (Burgers, interview Chicago, 24-08-2010). For guideline developers, evidence synthesis is not the primary goal, but the best *means* to reach their main goal: answering clinical questions. When Cochrane reviewers are faced with an absence of (high quality) evidence they may simply "conclude there is no conclusion" (Fieldnotes, Knowlege Translation summer institute, 23-06-2008), considering the identification of an absence of evidence a worthwhile achievement in itself. Guideline developers also see value in producing such 'known unknowns[7], but still feel compelled to look to other sources to answer their original question:

> "we haven't had a question yet that we haven't been able to answer. We haven't always had a great evidence base, but that doesn't preclude giving an answer."
> (Fieldnotes, Practice Guidelines summer institute, 23-06-2009)

Whether (and how) to provide recommendations in the absence of high quality Evidence is debated between guideline developers, but PEBC is expected to provide answers, even under conditions of uncertainty:

> *Guideline developer* : [a 'weak recommendation'] is an oxymoron: if you aren't sure, you shouldn't recommend something!

---

[7] The National Institute for Health and Clinical Excellence (NICE) in collaboration with several cancer centers set up a "database of cancer uncertainties (DoCU)" which collects and advertises the absence of evidence identified by guideline development, aiming to provide an agenda for further research: (http://www.nice.org.uk/aboutnice/howwework/researchanddevelopment/cocancpg.jsp)

> *Director of PEBC* : we have to say something, that is the whole point of making a guideline, a decision has to be made, even if it is based on an uncertain or weak recommendation. Our Ministry of Health demands an answer.
> (Fieldnotes, Practice Guidelines summer institute, 23-06-2009)

The need to have answers despite uncertainty is not a challenge unique to guideline developers – clinicians and policy makers alike do this frequently. The novelty (and difficulty) lies in PEBC's commitment to provide answers without evidence while simultaneously "being EBM". This was the great dilemma identified by EBM guideline developers at PEBC's symposium on the "Nature of Evidence" (Program in Evidence-Based Care, 2007).

Evidence Searched Guidelines

This seemingly contradictory question is answered with surprising confidence in the guideline development manual devised by the "Evidence Based Guideline Development" (EBRO) platform in the Netherlands:

> "Evidence-based means that a systematic search of evidence from literature has taken place and was reported on transparently. Nothing more, nothing less: if there was insufficient evidence – and this happens frequently – an answer still needs to be provided for the key question. Then the opinion of – and consensus between – the various experts will be decisive. We still call the method 'evidence-based': where possible, we have based our recommendations on evidence and made it explicit that the selected search methods did not provide any evidence."
> (EBRO, 2007: 4of80)

This conception of EBM guidelines - with or without Evidence - can be referred to as "Evidence Searched Guidelines", as they are not (necessarily)

based on Evidence, but differ from non-EBM (consensus) guidelines by the specification that consensus and opinion are only legitimate "*if* there was insufficient evidence". The systematic search is not only to find evidence, but to show that an absence of evidence reflects a 'truly unknown'. Even when little or no evidence is to be expected, much time, money and effort is devoted to the searching of evidence to *prove* such absence. And guidelines that are 'based on' evidence may not 'be EBM', if evidence that supported a pre-existing belief was selectively chosen and other evidence was ignored (Monaghan, 2008). The main EBM principle, then, is not the inclusion of high-level evidence per se, nor the exclusion of low level evidence or 'other' types of knowledge. The label 'Evidence-based' is reserved for guidelines that can show that only a specific *kind* of absence of evidence was allowed: it is compatible with evidence that is 'truly' absent, or known unknowns; and incompatible with evidence that is ignored, or intentionally unknown (Böschen et al., 2010: 786). The principles followed by Evidence Searched Guidelines reflect and respect the Evidence Hierarchy as the a systematic search is to ensure higher-level evidence is not ignored or rejected *in favor* of lower level evidence, opinion or consensus. This requires that evidence is ranked[8], including a clear indication when 'non-Evidence' (e.g. 'expert opinion') is included. The following sections will discuss the procedural rules that are to ensure guidelines follow the requirements of 'Evidence Searched Guidelines', by first addressing how the search defines the nature of Evidence, and secondly by distinguishing the kinds of 'non-Evidence' that are considered legitimate as a basis for guideline recommendations.

What counts as Evidence

The search technique that characterizes 'Evidence Searched Guidelines' is the same that is used for the production of systematic evidence reviews (see

[8] The assignment of Levels of Evidence is discussed in chapter three.

Moreira, 2007). Such a systematic search needs to find *all relevant evidence*, requiring both sensitivity (all the evidence) and specificity (only the relevant evidence), characteristics that are in tension with one another. To ensure that studies considered crucial by members of the GDG are located by the search, searches start very unspecific and hundreds or thousands of literature references are found. The main part of the search consists of evidence selection: distinguishing the evidence from the "noise" in which as little as 1% of references is retained as Evidence. This selection is not an automated fitting of medical literature into extant categories provided by the Evidence Hierarchy. First of all, the hierarchy lacks specific rules or cut-off points to distinguish Evidence from 'non-evidence', so each GDG agrees upon their own criteria to make this distinction. Secondly, the hierarchy does not provide any criteria for 'relevancy', and a number of different knowledge categories (biological, ethical and clinical) are needed to determine which references are relevant to the guideline's topic. To protect against arbitrariness in this situated and pragmatic selection process, the procedural rules of transparency and consistency are invoked.

What methodological designs count as Evidence?

Evidence hierarchies typically include 'expert opinion' or case studies as the lowest of Evidence, but guideline developers at CBO are instructed to agree upon a cut-off point for what they will treat as Evidence. Only 'Evidence' will be formally assessed by the methodologist and presented in Evidence Tables (see Moreira, 2007), while the 'non-evidentiary' literature may be described a 'qualitative' manner under the heading 'other considerations' (discussed below). In one group, participants had agreed that studies reporting on "a series of 5 patients or more" would be classified as Evidence ("observational study"), while reports on less than 5 patients would be categorized as "case studies" and not treated as Evidence. Yet, when the actual selection of medical literature was discussed, GDG

members did not always follow those selection criteria in a consistent way. An additional GDG meeting was called, not because participants could not agree on what Evidence to select, but to transform their pragmatic, informal and intuitive selection into consistent criteria that would fulfill the reporting requirements (GDG1d). GDG members criticized the arbitrariness of the formal selection criteria they had agreed upon, questioning how a study with five patients is *categorically* different from case studies:

> *GDG1member:* "What if you find 5 studies with 1 case? Does that count as a 'series of 5 patients' [and thus Evidence]?".
> [...]
> *GDG1chair:* We should be allowed to adjust the criteria. They are simply a choice that has been made.
> *Methodologist:* Yes, that's how it works. You've got your criteria, and then you adjust them. But you have to be transparent towards your readers.

The CBO methodologist, in charge of ensuring EBRO's procedural rules are followed, concedes that the definition of Evidence is somewhat arbitrary and adjustments are permitted for each GDG. There are no universal or validated cut-off points, but selection is considered 'systematic' by ensuring consistency throughout the guideline (for all GDG participants and all subquestions) and transparency (providing keysearch terms and selection criteria in appendices). While the criteria are somewhat arbitrary, consistency guards against randomness and 'selective' evidence selection ("cherry picking"). Transparency is to allow readers to assess the consistency of evidence selection. These procedural rules do not determine what evidence to select, nor do they nor simply 'reveal' a pre-existing process without change. Reporting the GDG's selection process with consistent criteria was a challenge that required an additional GDG meeting where GDG members reflexively reviewed their pragmatic evidence

selection process, and made changes both to the methodological inclusion criteria and to what evidence was selected.

Establishing relevance

Most of the literature references found by the systematic search will not be excluded because of flawed methodological design, but because studies report on diseases, patients or treatments that are not relevant for the guideline. The levels of evidence of the Evidence Hierarchy does not provide any criteria to assess the *relevance* of studies, leading to criticism that they over-value evidence that is of high methodological quality, but irrelevant (Grossman & MacKenzie, 2005; Cartwright, 2007; Dobrow et al., 2006). The design of randomized trials maximizes internal validity (i.e. the conclusions hold under the experimental conditions), at the cost of external validity (whether the conclusions are true in 'real life' is less certain). The validity of RCT evidence for 'real patients' has been questioned because experimental subjects are typically younger and healthier, and more often white and male than patients in the clinic (Epstein, 2007). Assessing the relevancy and applicability of Evidence is not only left to clinicians' expertise to properly "applying" evidence with patient's clinical circumstances and "values" (Sackett et al., 2000). The relevancy of (potential) Evidence is also assessed during guideline development. Determining what is 'relevant' or 'applicable' is a matter of establishing equivalences between the experimental conditions in which the evidence was produced, and the conditions in which the guideline is (imagined) to be used. GDG members not only have to decide *how* to apply evidence that is different in some way (Will, 2009), but have to agree whether the diseases, doctors, outcome measures, and health care arrangements are *similar enough* in the experimental and clinical conditions to count as Evidence in the first place. EBRO instructions urge GDG members to ensure that literature and guideline concern the same Patients, Interventions, Comparative treatment

and Outcomes (PICO) (EBRO, 2007: 17of80; EBRO, not dated: 6), and to ask whether “the results [are] applicable to the Dutch health care system” (EBRO, not dated: 9). EBRO does not provide any criteria or rules, only instructs that applicability “ought to be discussed within the GDG” (EBRO, not dated: 8).

On what basis do GDG members determine what differences do and do not matter? Catherine Will (2009) identified two general principles put forward by GDG members in assessing the relevance of studies when formulating a guideline: biological and ethical. My findings confirm these, and additionally identify the importance of knowledge of the current clinical standard to establish equivalences.

Pathophysiological knowledge

Firstly, to determine which differences between experiment and clinic are relevant, biological (or pathophysiological) mechanisms involved in the cause and/or treatment of the condition have to be assessed and compared. In some cases, differences in etiology were considered ‘irrelevant’, such that the experimental results from cancer patients were considered relevant evidence for patients with other diseases in the palliative phase. It was the similarities in symptoms (e.g. pain or depression) that made these studies relevant because the treatments under consideration were symptomatic treatments that did not target etiology. For the same reason, patients with the same disease may be considered incomparable: osteoporosis in women may have different causes from osteoporosis in men, affecting the effectiveness of (hormonal) treatments. Pathophysiological mechanisms are also important in establishing which studies concern relevant, and distinct, interventions. When assessing the effectiveness of treatment of depression, should studies on Tricyclic Antidepressants be included alongside Selective Serotonin Reuptake Inhibitors (SSRIs) as ‘anti-depressants’, or will studies on TCAs be excluded as ‘different’? Will all studies on the various SSRIs

(fluoxetine, sertraline, citalopram) be considered the 'same' intervention, or will some molecules (that may have more side-effects) be excluded as 'different'? (GDG2). When assessing analgesic treatment, will all studies on morphine be included as 'the same' intervention, or are studies on transdermally delivered morphine (patches) excluded, while oral and subcutaneous morphine are both included as 'Morphine treatment'? (GDG2).

When little is known about the etiology of the disease or the therapeutic mechanism of treatments (either in general, or by the guideline authors specifically), the question what is relevant Evidence becomes very difficult to answer. While such knowledge falls outside EBM's Evidence hierarchy, it is crucially important in determining what does and does not count as Evidence. To assure a reliable selection of 'all relevant Evidence', members of the GDG need to have sufficient pathophysiological knowledge of the condition, patients and interventions in question.

Weighing ethical principles and pragmatism to decide 'PICO'

Will (2009) discusses ethical principles governing clinical practice (e.g., accessibility and equity of care) as important in the application of Evidence. Ethical principles governing the production of knowledge (i.e., protection of trial subjects, equipoise, the immorality of withholding treatment) play a role in determining the nature of Evidence by playing a role in deciding the 'best' components of a study (i.e., the kinds of patients, interventions, comparisons and outcomes). Such ethical principles may reduce the relevance of Evidence, by deeming trials with children immoral, blinding in (sham) surgical interventions unethical, and demand that the current standard of care is upheld. In such cases ethics may trump methodological ideals of study design, and EBRO supports modification of the Evidence hierarchy 'in situations in which controlled trials are not possible because of ethical or other reasons' (EBRO, 2007: 19), positioning study designs other than RCTs

as the evidentiary gold standard. For example, when a guideline reviewer criticized the 'high quality' ranking of a study by questioning the relevance of its outcome measure, the guideline authors responded this evidence was 'gold standard' material because techniques change too quickly to continue trials with outdated treatments in order to measure longer term clinical outcomes (i.e. five-year survival) stating: 'It's the best Evidence we'll ever get. No need to wait for other Evidence.' (GDG5g). In other cases, the methodological selection criteria were modified not by ethical constraints, but by pragmatic constraints of Evidence production. In GDG4, an improved outcome for six months or longer was considered necessary to provide Evidence of long-term benefit. But since very few studies measured improvement for that long, it was reluctantly decided that all studies that measured improvement for three months or longer were counted as Evidence (GDG4), judging that more uncertain Evidence was preferable over less Evidence.

Knowledge of the clinic

Knowledge of the clinic is needed to determine what studies are relevant, which also includes organizational, financial and legal aspects of the current standard of clinical practice. Knowing the current standard of treatment is important to assess the relevance of comparative treatments. Trials may show a new treatment is more effective than placebo, an outdated treatment or a low dosage of the standard treatment. Such trials comparing to substandard care do not provide information whether the new treatment is better than the current standard of care. Guideline developers thus suggest to exclude such evidence as 'irrelevant', arguing that their aim is not to produce a knowledge synthesis, but to answer pressing clinical question: "we don't make academic publications, we make recommendations, so we don't look at evidence that doesn't help in recommendations." (PEBC guideline developer, interview 05-10-2009).

Knowledge about the availability of treatments may also be mobilized, such as when discussing a study about Diamorfine (heroin) as a treatment against pain. GDG members disagreed whether this study was to be selected as Evidence:

> *GDG2member1:* "You can throw that one [study on heroin] out, that's irrelevant, here in the Netherlands we can't prescribe that" (GDG2b).
> *GDG2member2:* No, this is important, it is about the effects of the opiate group as a whole. They are all opiates, with the same effects. […] We need to keep this study: it is recent, it is level of evidence A1, it is about older people, the same group really as the guideline. This is 1 of 3 of the same kind, 3 kind of opiates. I believe those are comparable.
> *GDG2Chair:* you believe that, or are you certain? [GDGmember1] says to throw it out…
> *GDG2member2:* It is one of a group, keep it in, we'll reformulate the recommendation… we shouldn't recommend this [heroin] treatment, but we should describe the study.

Knowledge of the clinic also includes assessing differences between the knowledge and skills of clinicians in practice and of those in the study, because clinicians' expertise may be an essential component of a (new) treatment, particularly in surgery. In one case, a surgeon's skills were considered such a distinct part of a new technique, that treatment results were deemed only applicable to that institution (GDG5g). They did not expect the same results from other surgeons conducting the new procedure, and did not want to recommend a treatment they considered dangerous when done by "inexperienced hands". Yet the study's methodological design was considered good and similarly designed studies had been included as Evidence. GDG members considered the same solution as above: including the study but without recommending the treatment. However, they found a

different criterion that disqualified it as Evidence. The study had not been located by the systematic search for evidence but its results had been presented at a conference attended by GDG members. Without a literature reference and lack of details, it could not be included and treated as Evidence.

Systematic search is not an automated search

In the logic of Evidence Searched Guidelines, the systematic search is to guarantee all (higher level) evidence is found and included, ensuring no Evidence is intentionally ignored in favor of lower ranked Evidence. A systematic search is not a mechanical search that operates independently from knowledge outside EBM's hierarchy. As we just saw, the search relies on the GDG members to specify the exact in- and exclusion criteria, and to agree on what diseases, treatments and patients constitute *relevant* evidence. The selection process is not determined by formal criteria of methodological design, but a range of 'other' knowledges, principles and categories are mobilized to determine what does and does not count as Evidence. The EBRO requirement of consistency and transparency does make 'cherry picking' more difficult, as the pragmatic and situated evidence selection needs to be justified towards the methodologist, other GDG members and the guideline reviewers and users. However, transparency does not mean the selection process is simply revealed unaltered and in full. The attempt to make things visible also produces blindness as "[e]ach standard and each category valorizes some point of view and silences another" (Bowker & Star, 2000: 6). Particularly since the formal criteria available for reporting are those of methodological design standardized by the Evidence Hierarchy. The lack of formal criteria to determine what is *relevant* Evidence means the important role that biological, ethical and clinical knowledge play in establishing the relevance of Evidence remains informal, unreported and invisible to those outside the GDG. Reporting the

evidence selection thus obscures rather than reveals the breadth of considerations and justifications relied upon in determining what counts as Evidence. However, in the concluding section I will discuss how the formal evidence criteria are diversifying by recent efforts to standardize and report relevancy criteria. But first I turn to the inclusion of 'non-Evidence' as a separate justification category.

Justifications that are not Evidence

Evidence Searched Guidelines allow guideline developers to claim an absence of Evidence as a 'true' absence. Then, it is legitimate to seek answers from "the opinion of – and consensus between – the various experts" (EBRO, 2007: 4of80). Such non-Evidence can inform guidelines only in the absence of Evidence, and to invoke this hierarchy non-evidentiary justifications are presented separate from - and secondary to - the category of Evidence, under the heading *Other Considerations*[9]. Such 'other considerations' can serve the same role as Evidence, but of a different (and lower) degree, providing a very literal space for non-Evidentiary knowledge such as 'opinion' as a basis for guidelines. Yet, its legitimate place is at the very *bottom* of the Evidence Hierarchy, never to trump Evidence, always secondary to it.

However, there is a second manner in which the non-Evidence included in 'other considerations' complements Evidence. Instead of an unintentional absence of evidence, the incompleteness of Evidence may refer to its indeterminacy. Evidence produced in very specific RCT contexts may not provide enough information to make practical and local recommendations :

---

[9] Other guideline organizations have categories of non-evidential justifications, such as "considered judgment" (SIGN in Scotland), 'Jugement Argumenté' (SOR in France), PEBC guidelines include such justifications in a less specific manner within the 'Discussion' section.

> "Clinical practice guidelines should be based on the best available evidence. However, this evidence is often incomplete, controversial, or lacking. Considerations beyond the evidence are therefore needed to be able to formulate *specific* and *applicable* recommendations for clinical practice." (Verkerk et al., 2006: 365 emphasis added).

When 'considerations beyond the evidence' are mobilized to supplement indeterminate Evidence, it differs from evidence in kind, not degree. Considerations needed to apply, translate and contextualize Evidence are not subordinate to Evidence, but *mediate*[10] between Evidence and recommendation, and in a mediatory role 'non-evidentiary justifications' can legitimately prevent recommendations being based on Evidence. The previously mentioned illegal status of heroin is one example of legitimate reasons why Evidence may be deemed 'inapplicable' to the guideline recommendation. The non-hierarchical relation is made possible by the lack of rules or mechanisms on how to weigh Evidence and "other considerations" in the formulation of recommendations. The great majority of guidelines report no "explicit link between the recommendations and the supporting evidence" (Graham et al., 2001: 159) and in practice the 'other considerations' can (and do) trump Evidence.

Some critics have argued that in order for 'other' kinds of knowledge - narratives, qualitative studies, clinical experience - to really count in Evidence Based Medicine, they need to be acknowledged as *kinds of evidence* and included in the formal Evidence Hierarchy (Lambert, 2006: 2641; Smith et al., 1996). The legitimate mediating role that "other considerations" play in guideline development suggest the contrary, lending support to Tonelli's (2006) suggestion to treat non-evidentiary justifications as outside the Evidence category, because they differ in *kind*, not degree.

---

[10] "Mediators transform, translate, distort, and modify the meaning or the elements they are supposed to carry." (Latour, 2007: 39).

Categorizing ‘other considerations’ *outside* the Evidence category does not mean they are ignored, but means they no longer need to compete with Evidence for priority. Allowing non-evidence to play a complementary role not only in the absence of evidence, but at all stages of guideline development, including as a mediator be between Evidence and recommendation.

The Nature of non-Evidence

The category ‘other considerations’ provides a literal space for non-evidentiary justifications in the development of EBM tools, but what exactly counts as legitimate ‘non-Evidence’? Little formal guidance on the nature of non-Evidence exists, the EBRO training manual mentions categories such as ‘patient preferences’, ‘cost considerations’ and ‘organization of care’ (EBRO, 2007: 74of80) but gives no explanation what types of information or documents to include within these general categories. This does not mean GDG members simply include any statement as a non-evidentiary justification. Rather, GDG members perform their own informal assessment to decide which claims, practices and expertise are deemed legitimate ‘considerations’ or ‘justifications’, and which are to be rejected as anecdotes, personal opinions or the institutionalized habits, or political or financial interests that the guideline is meant to change.

The acceptability of non-evidentiary justifications depends on processes of qualification, collectivization and/or standardization of claims. The existence of a textual reference of the source (an article or practice standard) also increases legitimacy. Four categories of legitimate ‘non-evidence’ can thus be distinguished: literature, qualified opinions, ethical principles and (practice) standards.

Non-evidentiary literature

The distinction between Evidence and 'other considerations' is not simply a distinction between medical literature on the one hand and expert opinion and consensus on the other. Non-evidentiary literature is literature that was not selected as 'Evidence' through the systematic search, because it was not located by it (conference abstracts, doctoral dissertations), or because it did not qualify as Evidence. For example, a commentary on a 'systematic' review was found to be very helpful, the GDGmember called it "a super-review. It's a review of a review article, more like an opinion on a Cochrane review, like a Super-Cochrane." The guideline methodologist suggest to include the Super-Cochrane under 'other considerations': "it may be retrieved [in full-text] and read to support and inform the working group's expert opinion, but it cannot serve as evidence" (GDG1d). This follows the EBRO manual that states: "it is wise to consult letters to the editor and/or editorials about the study you are assessing when you formulate your conclusions" (EBRO, not dated: 8). And such non-evidentiary literature can make a guideline 'more EBM', by improving the *critical appraisal* of Evidence (Will, 2010: 76-79).

Qualified opinions

Individual experts (within or outside the GDG) may provide information in order to justify and formulate a recommendation. But – even in absence of evidence - not all opinions of all experts are simply accepted. In one meeting, the members of GDG1 try to find the best 'expert opinion' by assessing the credentials of the opinion-holder (is s/he an author of a systematic literature review or practice guideline?) and evaluating what the basis is for their opinion (their own practical experience, their institutions' protocols, their doctoral dissertation?) (GDG1c). Not all 'experts' and all 'opinions' are considered equal, but experts who are considered international 'authorities' on the topic (authors of guidelines or literature

reviews) are favored, but so are local and junior researchers if they can support their opinions by referring to published literature.

Ethical principles

Thirdly, ethical guiding principles such as *primum non nocere* (first, do no harm) can provide rationales for (in)action: “from an ethical perspective it seems reasonable to [recommend intervention], primarily so as not to harm the patient” (GDG1 draft, page 113). Such ‘universal’ ethical principle can trump uncertain Evidence, as shown in the following exchange where an ethical principle was weighed against a single piece of Evidence:

> *GDG6member1:* If the [long-term] benefit of the medication hasn’t been proven, you have to stop [after five years]. That’s the principle of *In dubio abstine* [when in doubt, abstain].
> *GDG6member2:* it’s difficult. There’s no evidence for stopping the treatment either. For continuation, at least we have one study.
> *GDG6member1:* you can’t base anything on one study. Then we rely on *In dubio abstine.* (GDG6b)

GDGmembers 1 and 2 hold contrasting views of non-evidentiary considerations as secondary to Evidence (one study is better than no evidence), and ethical principles as outside the hierarchy all together. To trump such a universal principles takes *certainty*, not the provision of *a single piece* of (low level) Evidence. The (financial) accessibility of treatment is another principle that frequently trumps uncertain Evidence of the benefits or harms of a treatment (GDG2a,d).

Existing (practice) standards

Finally, important suggestions for guideline recommendations can come from existing standards for clinical practice or research (e.g. current clinical practice guidelines, RCT protocols or health care regulations). When

discussing what patient group was considered ‘high risk’ and should be screened, another guideline was suggested to provide guidance:

> *GDG6member1 :* the question is: “are there a lot of patients in this high risk group?”
>
> *GDG6member2 :* that depends on how we define ‘high risk’
>
> *GDG6member1 :* That is the NAPO [guideline] that has set that.
>
> *GDG6member2 :* well, that’s what you say, the NAPO. But we can also say “let’s define ‘high risk’ differently”. The NAPO is also simply a group of doctors gathered in a room that made up something nice. That is the reason for NAPO [‘s definition].
>
> (GDG6b)

Here the legitimacy of a guideline as a basis for recommendations is questioned, as any statement produced by their own group could provide a similar degree of justification. Yet *existing* guidelines making up the current ‘standard of care’ are not easily dismissed as *personal* opinions, routines or habits. Guideline developers frequently rely on existing practice standards and conventions to improve the practical use and do-ability of their recommendations:

> *GDG1member1:* The study by Coleman... That whole paragraph, doesn’t do it for me. It doesn't say *what* [pain medication] you have to give the patient.
>
> *GDG1member2:* There are pain guidelines, why don’t we just follow those?
>
> *GDG1Chair:* There’s no reason why pain management is different in [this disease].
>
> (GDG1c)

They accept this standard after (casually) considering whether there is any reason the current practice standard is not appropriate for this guideline.

Reasons to accept or reject the standard could invoke any of the aforementioned considerations of equivalency (biological or ethical principles, or knowledge of the clinic), between the existing practice standard and the guideline.

Like any type of standard, the nesting of standards is common and important when developing guidelines (Lampland & Star, 2009). Reporting standards set by journals, diagnostic standards set by international committees and (informal) ethical guidelines provided by professional training are all referred to in guidelines, reinforcing each other. Instead of considering the interdependency of standards a barrier to change - as innovations can only succeed by modifying an entire network of interconnected standards and routines - existing standards and networks can also as be used as a vehicle for change. New standards will function more easily by fitting them into the routines of an existing practice, instead of replacing them (Berg & Timmermans, 2000). By incorporating existing practice standards into a new standard, it may become an essential element within the existing interlocking network of standards that make up ‘local’ practice.

The Science of Evidence Based Guidelines

By focusing on the absence of Evidence this article has shown that knowledge outside the Evidence Hierarchy is not excluded from EBM tools. Non-Evidentiary knowledge and justifications are not only important in the application of Evidence in the clinic, but also in determining what counts as Evidence; and in formulating Evidence Based recommendations. What counts as Evidence is not an automatic application of formal classification criteria from the Evidence hierarchy. The features that define Evidence are very literally determined by selection criteria and cutoff points that have to be agreed upon by the members of each Guideline Development Group, who

mobilize a range of knowledges, principles and categories. And non-evidentiary considerations are not always secondary to Evidence; they also mediate between Evidence and recommendation, functioning independently from the hierarchy.

The kind of knowledge assemblage observed in guideline development shares features with the practical achievement of diagnostic classifications in the clinic; the categories of (genetic) science do not make clinical judgment and individualized knowledge redundant (Latimer et al., 2006; Rabeharisoa & Bourret, 2009). Both clinicians and guideline developers are aware of the contingent nature of their classification work, but differ in the way they "reflexively try to work around these contingencies" (Keating & Cambrosio, 2009: 340). For Evidence Searched Guidelines, procedural rules are to manage this contingency by the transparent reporting of consistent selection criteria agreed upon. While revealing some of the situated and pragmatic features of the evidence selection process, the procedural rules also obscure. EBRO addresses mostly criteria of study design, ignoring the process of assessing the relevance of evidence. The invisible nature of the assessment of relevancy may be changing, as it is undergoing formalization by a new procedural standard called GRADE.

Led by Gordon Guyatt (who coined the term EBM), GRADE is rapidly becoming the 'universal' standard for evidence assessment as it is used by dozens of guideline developers around the world. CBO has been involved in developing GRADE and is considering adopting it. GRADE proposes a more diverse Evidence Hierarchy, one not limited to methodological criteria ('study design'), but including assessment of a study's quality, the consistency of evidence, and a formal assessment of the relevance of evidence ('directness') (GRADE Working Group et al., 2004). GRADE formalizes some of the relevancy considerations that played an informal and invisible role in the GDGs I observed. This is not to say relevancy assessment is

automated. GRADE provides no fixed rules *how* to evaluate relevancy, only a checklist of factors to be considered by the GDG, the outcome to be reported on the standardized template of the GRADEpro software. This means the GDG members maintain much discretion of decisions: “the merit of GRADE is not that it eliminates judgments or disagreements about evidence and recommendations, but rather that it makes them transparent.” (Brozek et al., 2009: 669). GRADE claims to rationalize guidelines, not by excluding non-evidentiary considerations, but by revealing them for critical scrutiny by guideline users:

> “[GRADE] allows somebody else, who is taking the guideline and is applying it, to say: oohhh here is the basis on which they made the recommendation. If you follow their logic, I agree with their logic, therefore it is reasonable for me to follow the recommendation." (Guyatt, interview 7 October 2009, Hamilton).

EBM procedures like EBRO and GRADE, in order to find, assess and rely on Evidence are equally concerned with mobilizing, qualifying and reporting ‘non-Evidence’.

The regulatory objectivity of EBM guidelines

The rationalization of medicine that EBM promised in 1992 has been strongly criticized for ignoring the ‘subjective’ (but personalized and contextualized) expertise of individuals in favor of the ‘classical rationality’ of universal rules and quantified evidence (Cronje & Fullan, 2003). The findings presented here challenge the idea that EBM requires the elimination of experts’ judgments to be replaced by quantified evidence, but rather support the suggestion that the transformation brought about by EBM is “more than a simple regime change from rule by experts to rule by numbers” (Cambrosio et al., 2009: 655). This does not mean EBM has left medicine unchanged, or is ‘returning’ to a disciplinary objectivity, where the

judgments of experts is trusted by virtue of their professional status and/or moral virtues (Porter, 1995). The transparency central to EBM procedures signals a (perceived) mistrust of experts who are required to open up their actions, judgments and decisions to (external) auditing, standardization and scrutiny in an effort to *restore* trust (Blomgren & Sahlin, 2007: 166). The procedural rules described here are more compatible with the emergence of a new 'regulatory objectivity' as described by Cambrosio and colleagues (Cambrosio et al., 2006). Regulatory objectivity is based neither on quantified numbers nor the epistemic virtues of individuals, but invested in institutional procedures. Procedures regulating both what counts as Evidence and how (and which) experts are to act 'objectively'. This requires establishing a "deliberate and conscious formation of an internal consensus – often recognized to be pragmatic and situation-dependent – about how to proceed 'objectively'" (Cambrosio et al., 2009: 655). Moreover, the transparent reporting of the pragmatic consensus formation process serves to convince those who did not participate of the (regulatory) objectivity of that process.

These findings are in line with our earlier work, but provide a larger claim that procedural standards are central to the 'EBMness' of guideline development in general. Associating the nature of EBM with procedural standards that reveal the methodological process has implications for the four repertoires of evaluation that proposed by Moreira (2005). Moreira identified robustness (science), acceptability (politics), usability (practice) and methodological adequacy (process) as four kinds of justification relied upon in guideline development, but did not specify any interrelations between the four. The notion of Evidence Searched Guidelines indicates that EBM guidelines do not present a specific 'science' repertoire but a specific 'process' repertoire: "The term 'evidence-based guideline' mainly refers to the methodology [of guideline development], and not the robustness of the recommendations" (EBRO manual, 2007: 4of80). The methodological

process repertoire can function as a 'meta-repertoire'; providing rules how to perform the other repertoires. The EBRO and GRADE procedures that characterize EBM guidelines are largely concerned with the robustness repertoire (how to search, select, assess and present evidence; how to manage non-Evidence). Chapter five will introduce some of the international collaborations supported by the Guidelines International Network that produce methodological standards for other repertoires: to improve acceptability (e.g. external and peer review); usability (e.g. pilot-testing, multidisciplinary GDG membership); and the process repertoire itself (e.g. declarations of conflicts of interests) (Knaapen, 2012).

Regulation by guidelines for guidelines

The transparency required of the production of EBM tools suggests a new approach to EBM's "technologies for transparency" that increasingly organize and regulate health care (Blomgren & Sahlin, 2007: 161). Previous literature mainly examined transparency as an effect of the use of EBM tools that monitor and evaluate medical practitioner's work. The visibility produced by standards allows for external (or intra-professional) quality evaluation and control, resisted by professionals who consider it a threat to their autonomy, skills, discretion and trust (Power, 2000; McGivern & Fischer, 2012; Armstrong, 2002). Empirical studies examining the *kinds* of visibility produced by medical standards have found transparency and audit can also enhance professional's autonomy, by controlling *what* is made measurable, comparable and governable (Levay & Waks, 2009; Berg et al., 2000; Bowker & Star, 2000; Castel, 2009). Central to understanding the regulatory impact of any transparency tools is the question 'who controls quality control?' This question entails not only who controls how auditing tools are *used* (Power, 2000), but more importantly by whom and how such tools are designed and *produced*.

The producers of EBM guidelines aim to increase legitimacy and create trust by making their own practices transparent, expecting guideline users to hold them accountable for the 'objectivity' of their evidence selection and expert judgments. Putting the quality control of guidelines (and the 'objectivity' of those developing them) in the hands of guideline users suggests a type of mutual intra-regulation between users and producers of guidelines, avoiding an endless regression of controlling the quality of quality control. This assertion suggests the need for more critical research in at least two areas. First, transparency requires "a transparency-literate public" that is willing and able to understand and act upon what is made visible (Garsten & Lindh de Montoya, 2008: 13). Although relieved from the burden to critically assess medical literature themselves (Guyatt et al., 2002), clinicians are now expected to critically assess guidelines and all its appendices to evaluate whether procedural requirements ensuring the (regulatory) objectivity of the guideline have been met. Whether clinicians have the time, skills and tools to do so remains an open question. Secondly, guideline users monitoring guideline producers does raise the question of who controls *what* is to be revealed to guideline users? How is the quality and adequacy of 'guidelines for guidelines' (and the 'objectivity' of *their* producers) ensured? The new guideline 'scientists' collaborating with GIN to develop standardized procedures that require transparent guideline development (e.g. GRADE, AGREE), hail from different professions (epidemiologists, GPs, librarians, info specialists) and assert unique expertise as independent experts in guideline methodology (Knaapen, 2012). The meta-standards designed, promoted and enforced by these guideline methodologists fits neither the model of intra-professional regulation by medical professionals, nor that of external control by public institutions or (private) managers, but suggest the emergence of new meta-regulatory institutions and individuals that are specifically devoted to 'quality

control of quality control' in medicine. This international regulatory network is the central focus of the next chapter

CHAPTER 5

European regulation and harmonization of Clinical Practice Guidelines

This chapter first appeared in: *European Union public health policy: regional and global trends*. S.Greer & P.Kurzer (eds) Routledge, 2012: 64-80

Countless standards do nothing. Some, however, obtain majestic results.
(Timmermans & Epstein, 2010: 81)

European governance and Clinical Practice Guidelines

The European Union (EU) has few legal tools and administrative institutions at their disposition to regulate health care systems in Europe, as this domain primarily remains the legal mandate of national governments. Instead of a traditional command-and-control-type government, the EU has developed policies to encourage various actors (non-state, public, private, international, NGO's) to form 'networks' that establish their own standards and diffuse best practice (Hervey 2008). The term "New Governance" is often used to refer to the broad range of regulatory tools, models and techniques to describe the focus away from the EU as an administrative institution that uses legislation to govern, towards 'practices of governance', or 'multi-level' governance (Delanty & Rumford 2005: 139; Scott & Trubek 2002). New governance is expected to expand EU's regulatory impact "by utilizing new forms of knowledge and making use of global networking", instead of legal enforcement by a central authority (Rumford, 2002: 72). The specifics of such 'voluntary' mechanisms vary, but 'policy learning' and the production and exchange of comparative information (e.g. benchmarking, name and

shame) are considered viable strategies. EU officials and scholars alike have high hopes for such new governance structures to tackle the challenges of contemporary health care systems (Trubek et al., 2008). The EU increasingly invests time, money and hopes into the development of standards, data and networks to regulate health care professionals and promote better health care practices in Europe (Greer, 2006), but few studies have empirically investigated the impact of specific European quality or safety standards in the health domain (Hoeyer, 2010). Without traditional and centralized modes to impose changes on actors many scholars and EU policy makers remain skeptical about the mechanisms by which voluntary standards, knowledge and networks can achieve effective regulatory impact. This chapter will present an empirical study of the establishment of the Guidelines International Network (GIN) to demonstrate the specifics of how EU funding was successful in establishing a trans-national governance structure that regulates through knowledge and voluntary standards, yet did not succeed in a harmonization of European health care practices.

Clinical Practice Guidelines (CPGs) are a particular type of medical standard that aims to regulate physicians' behavior through knowledge. Based on a synthesis of evidence from medical research (ideally randomized clinical trials), CPGs provide clinicians with recommendations how best to treat a particular clinical condition. By providing a Gold Standard for clinicians to 'live up to', such evidence-based guidelines carry the promise of simultaneously rationalizing and standardizing medical practices (Timmermans & Berg, 2003). Since the late 1990s the European Union has invested in the development and harmonization of Clinical Practice Guidelines by funding several closely connected research projects in this domain (Table 1).

Founding year	1998	1998	2001	2002	2005	2006
Nationality	AGREE	GRADE	European Council Recom	GIN	ADAPTE	CoCanCPG
Canada	Brouwers				Brouwers	
	Browman				Browman	
Finland	Mäkelä		Mäkelä	Mäkelä		
France	Fervers			Fervers	Fervers	Fervers
Germany	Ollenschläger	Ollenschläger	Ollenschläger	Ollenschläger		
Netherlands	Burgers			Burgers	Burgers	
	Grol					
Scotland	Qureshi			Qureshi		Qureshi
Switzerland	Burnand	Burnand		Burnand	Burnand	
UK	Cluzeau	Cluzeau		Cluzeau		
	Feder					
	Littlejohns			Littlejohns		Littlejohns
US	Slutsky			Slutsky		

Table 1. The AGREE Collaboration led to many international guideline projects.

These projects have resulted in a quality standard for guidelines (AGREE), a formal recommendation on guideline methodology by the European Health Committee (Council of Europe, 2002) and the establishment of the Guidelines International Network (GIN). This chapter will chronicle how these EU funded projects have been successful in establishing a sociotechnical network whose voluntary standards have widespread regulatory influence in the guideline 'world'. The projects were successful by realizing a change in regulatory structure : from centralized European guidelines to a distributed knowledge network. But this also meant transforming *what* was being regulated: from standardization of guidelines (products) to standardization of methodology (process). The result was a network that rejects the production of centralized European guidelines with which to harmonize health care practices in Europe. If EU policies pursue the establishment of networks as a policy goal in itself, with little attempt to coordinate harmonization policy with networks, European New Governance may be highly effective, but the EU will have little control over *what* it is effective at.

The regulatory power of standards, knowledge & networks

To better understand the technical and knowledge practices that are central to giving voluntary standards regulatory impact, this chapter draws upon the field of Science and Technology Studies (STS). While 'new governance' may be relatively new as an explicit and formal EU policy, STS scholars have repeatedly highlighted the importance of standards and standardization to regulate almost all aspects of modern life (Timmermans & Epstein, 2010; Busch, 2011). STS work provides myriad examples of successful and transformative (voluntary) standardization driven by practices of knowledge production. Scientific methods and classifications are not 'naturally' and automatically universal, but are "…a triumph of human organization - of regulation, education, manufacturing, and method." (Porter, 1995: 29). For 'universal' facts and artifacts to travel across boundaries of time, space, professional groups or languages myriad standardization processes are required (O'Connell, 1993; Collins, [1985]1992). By emphasizing the "leaky borders" between scientific categories, technical infrastructure, social norms and institutional routines or personal habits (Lampland & Star, 2009) scholars also emphasize the normative nature and political consequences of standards, classifications and knowledge (Bowker & Star, 2000). The co-production of standards and knowledge takes place within a sociotechnical network in which both technical infrastructure (databases, standardized terminology, calibrated apparatus) and 'social' connections (personal contacts between colleagues, peer review and peer pressure, tacit knowledge and face-to-face learning) are indispensable. In order to capture the mechanisms that give such sociotechnical networks regulatory 'power', requires a departure from traditional conceptions of agency. Rather than including only humans as agents of change, STS has generally come to accept that all the interacting elements that constitute a sociotechnical

networks are “actors”. Neither humans nor objects by themselves effect change, it is the strength of the connections between all constitutive elements that give a network regulatory and normative power (Latour, 1983; Callon, 1986). Objects and tools are not “merely passive objects of human manipulation” but as constitutive elements in a network they (co-)produce effects, they *do, perform* and *change* things (Barry, 2001: 11). Although people (such as those listed in Table 1) are indispensable to this story, this chapter is not only about their ambitions, interests, beliefs and intentions, but objects (quality standards, evaluative instruments, comparative data) are equally important vectors of action.

This chapter draws on empirical material that was collected by studying international collaborations concerned with guideline development that received EU funding (Table 1). Besides document analysis, I conducted semi-structured interviews with five founding members of these projects; conducted participant observation (as user and evaluator of the AGREE instrument); attended international working group meetings (ADAPTE, G-I-N PUBLIC) and the annual conferences of the Guidelines International Network (2007-2011). Drawing on this empirical material, the chapter chronicles the emergence of a new sociotechnical network concerned with international regulation of clinical practice guidelines. I first introduce Clinical Practice Guidelines and the promises and challenges they posed to the standardization of medical practices, setting the stage for the European Union’s interest in funding collaborations to harmonize such guidelines. These efforts reject the standardization of guidelines on a European or international level, instead proposing the harmonization of the *methodology* to develop guidelines. It was the AGREE instrument, which, by defining the quality of guidelines in terms of its methodological process, transformed international guideline development. This chapter will discuss three different aspects of this transformation. Firstly, it modified the notions of standardization and rationalization of Evidence Based Medicine. Secondly,

by producing comparative knowledge it standardized guideline development practices around the world. Thirdly, by founding the Guidelines International Network (GIN) it established a regulatory infrastructure that seeks to establish experts in guideline methodology as a new self-regulating profession. The chapter concludes with the claim that EU funding has been successful in establishing a 'new governance' structure whose distributed regulatory impact has simultaneously reached beyond the EU, and is more localized than a nation state. The standardization of guideline methodology has professionalized this domain, establishing a novel meta-regulatory arrangement in medicine. Nevertheless, far from Europeanizing guidelines, this international network entrenches the development of guidelines further in national organizations.

The promise of guidelines as Gold Standards

Clinical practice guidelines are a specific kind of medical standard which are most commonly defined as: "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances" (Field & Lohr, 1990: 8). They first emerged in the 1970s, when the widespread concern about the efficacy and cost of medical practice was heightened when analysis of treatment patterns in the USA showed that patients with the same condition were treated differently depending on geographical region. The production of guidelines was offered as one of the solutions to end unwarranted variation. They could function as 'gold standards' that recommend and disseminate best practice to physicians in the clinic (Weisz et al., 2007). In 1990s the Evidence Based Medicine (EBM) paradigm emerged with the aim to "use the literature more effectively in guiding medical practice" (Evidence Based Medicine working Group, 1992: 2024). They emphasized that the fallibility of 'intuition and unsystematic clinical experience' applied not only to local practitioners but

also to national experts. They emphasize guidelines are “only worth following” when based on evidence from medical literature, especially randomized clinical trials (RCTs) (Sackett & Rosenberg, 1995). By shifting the determination of “best practice” from individual clinicians, to national experts, to quantitative evidence, guidelines followed a more general shift in the ‘scientific’ basis of medicine: “from a regime of trust in expertise and experts to a regime based on the mechanical generation of data, the elimination of human judgement and the adoption of a ‘view from nowhere’” (Keating & Cambrosio, 2009). Cronje and Fullan see EBM as an attempt to model medicine on the “classical rationality” of mathematics and logic: “The appeal of the classical model of rationality in medicine, operationalized by EBM, is clear: it promises a rule-governed procedure that, if followed faithfully, will necessarily result in improved health outcomes for all patients.” (2003: 355). In this regime rationalization is identical with standardization: “any rational person, if s/he begins with the same information, will arrive at the same conclusions” (2003: 355). The appeal of guidelines is clear, by providing the evidence and rules to follow, they will automatically result in better medicine. With such high hopes, thousands of ‘evidence based’ guidelines for clinical practice are produced by private, professional, public and governmental organizations. For the European Union, guidelines promise a way to harmonize European medical practices without recourse to health care legislation or regulation that is the legal competence of national states. In the late 1990s the EU starts funding international collaborations of guideline developers to harmonize guideline development in Europe (Table 1).

European regulation of guidelines

The first survey of clinical practice guidelines in Europe found that the success and proliferation of guidelines posed a new challenge. It stated that the diversity in guideline-setting initiatives in Europe presented a “danger

that in some years healthcare providers will drown in a guideline morass." (Grol et al., 1998: 65). The "overdose" of conflicting guidelines of varying quality questions the vision of guidelines as rational Gold Standards. Guidelines appear to reflect the practice variation they were meant to reduce, and are themselves in need of standardization and rationalization. In 1998 a consortium of guideline researchers is funded by the EU to develop the AGREE instrument, an evaluative instrument to address "the number of guidelines in specific clinical areas that contain conflicting recommendations" (Cluzeau, 1998), and two of AGREE's members contribute to the European Health Committee's recommendation on guideline development addressing similar concerns (Council of Europe, 2002: 12). AGREE members continue to establish or participate in a range of international collaborations that are supported by EU funding such as ADAPTE, GRADE and GIN (Table 1). The AGREE instrument comes to underpin international regulation of guidelines by sharing a distinct vision of international collaboration.

The official vision for international collaboration on guidelines articulated by the self-acclaimed 'father' (Jako Burgers) and 'mother' (Françoise Cluzeau) of the AGREE instrument rejects the discourse of guidelines as Gold Standards based on quantitative evidence that will standardize medical practices. They emphasize that in addition to evidence "expert opinions, practitioners' and patients' preferences as well as societal priorities" play a role in standard setting and conclude "each country has its own norms and values that influence the content and presentation of guidelines. Therefore, the aim should not be to develop international guidelines" (Grol, Cluzeau & Burgers, 2003: S6). The European Health Committee's Recommendation on guidelines reflects the same concerns and instead of proposing a single 'best' Gold Standard for Europe, it emphasizes the necessity of national and local diversity in guidelines in order to "suit the practical circumstances of the organisation applying it", "to

accommodate the specific needs of guideline users", and because "implementation is enhanced" (Council of Europe, 2002). As the EU subsequently funds other projects that reject the international standardization of Clinical Practice Guidelines (Table 1), the promise of guidelines to regulate physicians directly with 'universal evidence' – avoiding national legislation or regulation – does not materialize and guidelines remain national documents.

From European Guidelines to European guideline development

The rejection of guidelines as universal (or European) documents that can standardize medical practices is not simply a resistance to standardization or globalization. International collaboration on guidelines does not end, but shifts towards reaching "international agreement about the requirements for methodology and reporting of guidelines." (Grol, Cluzeau & Burgers 2003: S6). Harmonization of the *process* of guideline development becomes the focus of European policy as expressed in recommendation Rec(201)13 "*on developing a methodology for drawing up guidelines on best medical practices*" (Council of Europe, 2002). Increased international networking between "organisations, research institutions, clearing houses and other agencies that are producing evidence-based medical information" was also recommended (2002: 15), and in 2002 the Guidelines International Network (GIN) was established with EU funds. GIN now includes guideline development organizations from 46 countries, and around 400 guideline developers attend its annual meeting to exchange information, learn methodology and collectively define, defend, disseminate and enforce 'good' guideline development. By establishing GIN, EU funding contributed to the creation of a network that creates, promotes and disseminates *the* international methodological standards for guideline development to which hundreds of guideline developers submit voluntarily. Significantly, not only the regulatory structure changed (from centralized European standards to

the distributed regulatory power of a network), but *what* was being regulated had also changed: from standardization of guidelines (products) to standardization of methodology (process). It was the AGREE collaboration that was instrumental in achieving this shift.

AGREE: guideline for guideline development

In 1998 an international collaboration of guideline developers from predominantly European countries was funded under the BIOMED2 program of the European Commission to develop the *European Union Critical Appraisal Instrument for Guidelines* (EUCAIG), "an appraisal instrument to compare the different approaches to guideline development in Europe". It was renamed the catchier AGREE; *Appraisal of Guidelines Research & Evaluation in Europe*, and aspiring to be the worldwide standard, eventually dropped "Europe" from its name. The AGREE instrument consists of a checklist of 23 items in 6 domains to assess whether "rigorous" guideline development methodology has been followed in the production of the guideline being appraised. It asks questions such as whether a systematic literature search has been done, if various 'stakeholders' are included and if the guideline will be updated regularly. AGREE thus evaluates guidelines based on the quality of (the reporting of) the methodological process, and "does not assess the clinical content of the guideline nor the quality of evidence that underpins the recommendations" (AGREE Collaborative Group, 2003). Some consider this a serious limitation of AGREE (Vlayen et al., 2005; Hannes et al., 2005), as methodological quality only loosely correlates to measures of the quality, acceptability or validity of the content of the guideline (Watine et al., 2006; Nuckols et al., 2008). The founder members of AGREE, such as Jako Burgers, acknowledge that an AGREE score cannot *guarantee* the clinical validity of a guideline, but find it reasonable to assume 'rigorous' methodology makes a good guideline more

“probable”. By jokingly adding “I think it has an odds ratio of 2 out of 3” Burgers acknowledges but gently mocks the need to confirm assumptions with formal statistics, as the Evidence Based Medicine paradigm customarily requires (Burgers, interview 24-08-2010, Chicago). In practice, AGREE scores are treated as a pragmatic shorthand for, or a straightforward representation of, the quality and validity of a guideline (Burgers, Cluzeau, Hanna, Hunt & Grol, 2003; MacDermid, Brooks, Solway, Switzer-McIntyre, Brosseau & Graham, 2005). The AGREE collaboration has thus positioned ‘rigorous guideline methodology’ - not the quality of evidence or the clinical content of guidelines - as the primary ground for evaluating the validity of guidelines. Disconnecting the quality improvement of guidelines from the standardization of guidelines allows guidelines to vary from country to country, yet be legitimate if the process of arriving at the guideline is ‘up to’ standard.

Transforming the debates of Evidence Based Guidelines

Rejecting global standards for failing to capture the complexities of (local) culture, values and environment is not new. It is a typical narrative that contests standardization for bringing about a dehumanizing uniformity that fails to capture the complexities that matter ‘on the ground’ (Timmermans & Epstein, 2010: 71). Guidelines have been at the center of similar contestations over whether Evidence Based Medicine brings about rationalization or standardization. Medical professionals have criticized clinical practice guidelines for creating “cookbook” medicine that ignores the expertise of physicians and values of patients (Berg, 1997; Lambert, 2006). The limitations of ‘universal’ quantified evidence to properly capture the complexity of decision making in clinical practice have also repeatedly been pointed out (Goldenberg, 2009). Many critics suggest a stark divide between

"EBM advocates" who strive for "rational medicine" (which is universal) and those who resist EBM and strive for a "humane medicine" by preserving the "clinical art" (Timmermans & Mauk, 2005; Mykhalovski & Weir, 2004). The international regulation of guidelines is remarkable as it fits neither side of such a polarized debate. They acknowledge the limitations of quantified evidence and problems of standardized medicine, but instead of rejecting guidelines and EBM they offer 'systematic' guideline methodology as a solution to this problem. This new justification for guidelines proposes a new rationality for guidelines that cuts across both sides of the debate. It allows guideline developers to defend evidence-based guidelines while avoiding the unrealistic (and inaccurate) ideal of guidelines as Gold Standards.

We have already seen guideline developers collaborating in AGREE reject "classical rationality" as variation in guidelines is legitimized. Quantitative evidence such as RCTs and systematic reviews are important, but they do not *automatically* rationalize medical practice, only after guideline developers have conducted a process in which evidence is found, translated, interpreted and implemented for the specific context (Grol, Cluzeau & Burgers, 2003: S5-6). Instead of simply trusting individual guideline developers to perform this evidence translation, international guideline collaborations propose rigorous' and 'systematic' guideline development procedures to ensure the rationality of evidence contextualization. AGREE's 23 items provide requirements for assessing quantitative evidence as well as ensuring the presence and 'objectivity' of 'local' experts and stakeholders. The outcome of the process is neither

universal nor based on the opinion of *individual* clinicians or experts, but specific to the practice in which the guideline will be used[11].

The GRADE working group is one such international guideline collaboration that receives EU funding to rationalize the translation of evidence (Guyatt et al., 2008). GRADE is led by Gordon Guyatt of McMaster University who coined the term Evidence Based Medicine, and includes three AGREE founding members. It attempts to standardize the quintessential (and much criticized) EBM tool that assigns quality to evidence: the 'evidence hierarchy'. Despite terminological standardization and the use of the same software around the world, GRADE has not been able to establish universal rules how to translate evidence. GRADE developers consider it a success because by revealing (rather than eliminating) judgments, guideline users can scrutinize (rather than simply trust) expert judgments. Standardized reporting of the guideline developmental process:

> "allows somebody else, who is taking the guideline and is applying it, to say: oohhh here is the basis on which they made the recommendation. If you follow their logic, I agree with their logic, therefore it is reasonable for me to follow the recommendation" (Guyatt, interview 7-10-2009, Hamilton).

---

[11] The validity of 'universal' evidence or standards for 'local' settings is a familiar and contested issue in most standardization processes. The 'contextualization' of evidence during standard development is only one of many possible solutions to bridge the radically different environments of 'context-free' RCTs and clinical practice. Timmermans and Berg (1997) describe how 'universal' standards require informal work-arounds and local adjustments to be workable in practice. To avoid this 'implementation gap', Zuiderent-Jerak (2007) suggests standards to be produced *within* the practice where they are to be used, which he calls situated standardization. Carl May (2006) describes how 'universal' quantified evidence was rejected outright in favor of what he calls "practice based evidence" (local experimentation and qualitative evidence). Others advocate for prioritization of evidence from observational studies (Black, 1996), which would rely on entirely different kinds European networks, such as EUROCARE (see Briatte, 2012). Attempts to 'contextualize' or localize the production of randomized clinical trials include what Epstein (2007) calls "niche standardization", as well as comparative effectiveness research (Keating & Cambrosio, 2012: 376-381).

Instead of a mechanical view of guidelines in which evidence determines "correct judgments" and single outcomes, international guideline developers claim the evidence-based approach is defined by a methodological process that can "explicitly represent the issues for competing arguments and foster critical thinking by insisting on accountability to evidence." (Brouwers, Somerfield & Browman, 2008: 1026).

Redefining EBM's rationality based on a "rigorous" and transparent methodological process cuts across the polarized debates that surround EBM. This new legitimacy allows guideline developers to respond to the critiques that Evidence Based Guidelines encounter. The legitimacy of diversity of guidelines and judgments reduces fears that guidelines will erase distinct cultures, human values and clinical expertise necessary for high quality medicine. The insistence on the 'contextualization' of evidence reduces concerns of EBM's over-reliance on quantification and reductionism. By promoting universal methodological procedures and local guidelines; emphasizing the need for universal evidence and contextual judgments, allows guideline developers to claim their guidelines are universal *and* local; are objective *and* include values. As Lambert (2006) has noted for Evidence Based Medicine in general, the limitations and critiques launched at EBM are not overcome by rejection or contestation but become incorporated. By focusing on process standardization instead of product standardization, AGREE has turned the variation in guidelines into a resource for more legitimate guideline development.

AGREE standardizes guideline development

AGREE not only provides an abstract ideal for guideline developers, it is a sociotechnical instrument whose normative impact transforms and regulates the work of guideline developers around the world. And while AGREE was

not the first to formulate methodological standards for guideline developers (the highly regarded Institute of Medicine in the USA had outlined similar ideals a few years earlier (Lohr & Field, 1992)) it is AGREE that has had significant standardizing impact around the world. AGREE's regulatory impact is not the result of a more authoritative institutional origin, or the imposition of sanctions. AGREE distinguishes itself because it not only expresses an ideal to live up to, but provides a test of that ideal (Busch, 2011: 52). Unlike other methodological standards, AGREE provides an evaluative checklist that produces a numerical, quantified score of guideline quality (Vlayen et al., 2005).

Many guidelines assessed by AGREE receive what seem to be rather low quality scores, for example 85 per cent of a collection of Canadian guidelines scored less than 5 out of 10 for 'rigor of guideline development' (Graham et al., 2001). For fear of "misuse" of their instrument, AGREE never set a formal quantified norm or cutoff point (Cluzeau, interview 3-11-2009, Lisbon), so AGREE's numerical scores *alone* do not provide a basis to reject or approve guidelines. But by quantifying quality, AGREE's quality measurement is no longer "alone", but easily embedded, transferred, circulated and compared. The numerical score makes comparison and ranking of guidelines with entirely different content, format, clinical topic, produced by different institutions, languages or countries possible. By ranking scores, over time or between guidelines, a relative norm is created. The act of quantifying produces comparative knowledge with normative effects (Espeland & Stevens, 2008). In the medical literature guidelines and guideline programs receive negative publicity because they scored 'low' on the AGREE, i.e. lower than others (Brouwers & Charette, 2001), guideline development organizations are encouraged to reform their developmental programs to increase their AGREE scores in the future (Hurdowar et al., 2007: 657); and in the USA AGREE scores have been

used to discredit guidelines that insurance companies relied upon to deny coverage (Manchikanti et al., 2008).

The normative pressures provided by AGREE are amplified by requirements that EU funded projects use AGREE (Cluzeau, interview 3-11-2009, Lisbon). For example, the Coordination of Cancer Clinical Practice Guidelines (CoCanCPG) is a EU funded project that has established cooperation between 17 cancer guideline programs in and beyond Europe (Fervers et al., 2008) evaluated whether cancer guideline developers comply with the AGREE criteria. And although the comparative knowledge created by this project is 'descriptive', it creates a normative picture showing guideline programs 'where they stand' on the world ranking, creating normative pressure for 'non-compliant' programs to adjust to the international gold standard set by AGREE (Figure 2).

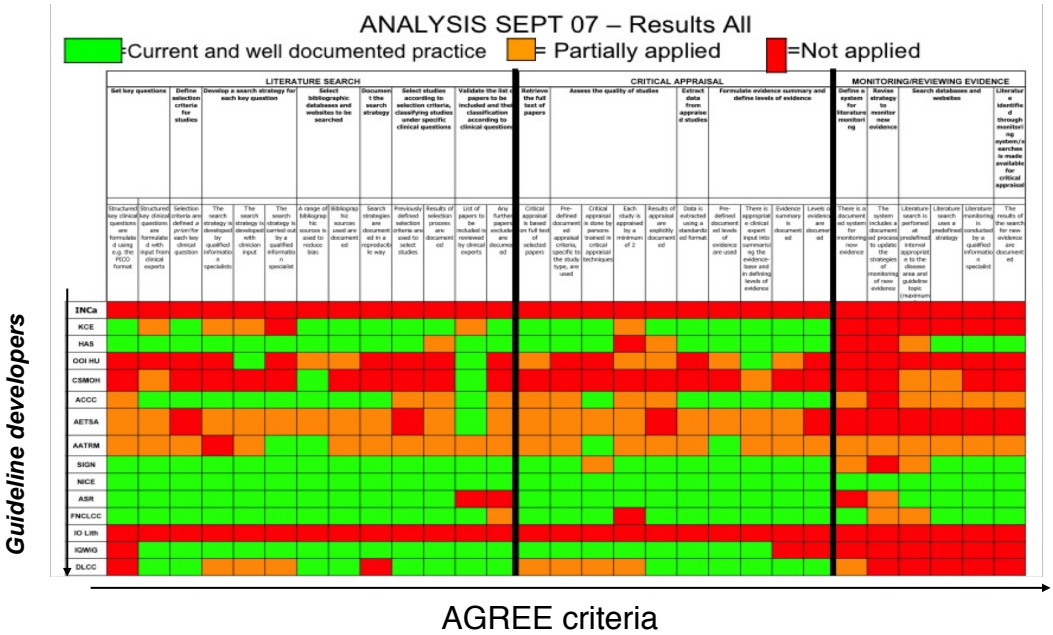


Figure 2. Benchmarking guideline development programs (vertical axis) along AGREE criteria (horizontal axis). Modified from: B. Fervers, "Benchmarking guideline development programs", presentation at DLCC Montreal, October 9th 2007 (slide 1)

AGREE's status as the gold standard of guideline development has become uncontested around the world. Its founding article has been cited 324 times (22-02-2012 Web of Science), it has been evaluated as the most useful of 24 guideline appraisal instruments (Vlayen et al., 2005). More than 1100 oncology guidelines, including those from the UK, the Netherlands and European medical societies, have been evaluated with AGREE, the scores easily available in an online database, making comparison of AGREE scores very easy for physicians, policy-makers and the public (www.cancerguidelines.ca). As guideline developers anticipate their guidelines may be evaluated with AGREE in the future, the AGREE checklist is not only used to evaluate guidelines, but has become a 'manual' on how to develop a guideline (e.g. Bloem et al., 2010: B2), and how to set up a guideline program (Van der Wees et al., 2007).

The AGREE instrument is a 'voluntary' quality standard that simultaneously is an ideal to live up to, an *evaluative* instrument measuring the state of guideline methodology, and a normative standard that transforms and performs guideline methodology. It expresses and measures the ideal of guideline development, and by providing numerical scores that allow easy comparison of guidelines and guideline programs, it makes (relative) quality or its lack thereof visible, persuading guideline developers to modify their programs towards the ideal embedded in the AGREE instrument. AGREE has not only "*demonstrated* an increasing harmonization of the methodologies used by guideline agencies and programs around the world" (Ollenschläger et al., 2004: 456, emphasis added), it has been instrumental in *producing* this standardization.

The Guidelines International Network

The AGREE collaboration solidified its standardization impact through the establishment of the Guidelines International Network (GIN) which becomes

a single platform to promote, connect and embed the standards and knowledge produced by projects such as AGREE, CoCanCPG and GRADE. GIN was founded as a result of the AGREE collaboration's renewed funding under the 5th Framework whose objective was to promote AGREE and establish "an international network of guideline agencies and related organizations". Previous research conducted by AGREE members had shown large divergence in guideline development initiatives throughout Europe, both in the content of guidelines (Burgers et al., 2002) and in the methods of guideline development (AGREE Collaborative Group, 2000). Much guideline development takes place at organizations for which guideline development is only a minor activity, such as medical specialty organizations, regional health care networks, academic institutions or health care financers. Individual guideline developers may be information specialists, librarians, epidemiologists or medical professionals and do not share national professional associations, educational trajectories or even job titles. So while the AGREE collaboration brought together a dozen guideline developers from large national organizations exclusively dedicated to guideline development such as NICE (UK), CBO (Netherlands) and AZQ (Germany), GIN does more than providing an international platform for guideline developers previously organized on a national level[12]. GIN brings together 'professional' guideline developers that previously did not share any identity or community on a national, regional or even local level. The knowledge about the diversity in guideline development in Europe served to make the contours of a distinct (albeit diverse) 'guideline world' visible for the first time, providing a picture of who is participating, in what way, and what might be improved. The results of one such survey were presented at an

[12] In the Netherlands a similar informal network of guideline developers call 'EBRO platform' did exist. As neither organization nor individual, the EBRO Platform did not become a GIN member, but several of EBRO's members did.

international workshop on guideline development to advertise the need and desirability of an international guideline network (Figure 3).

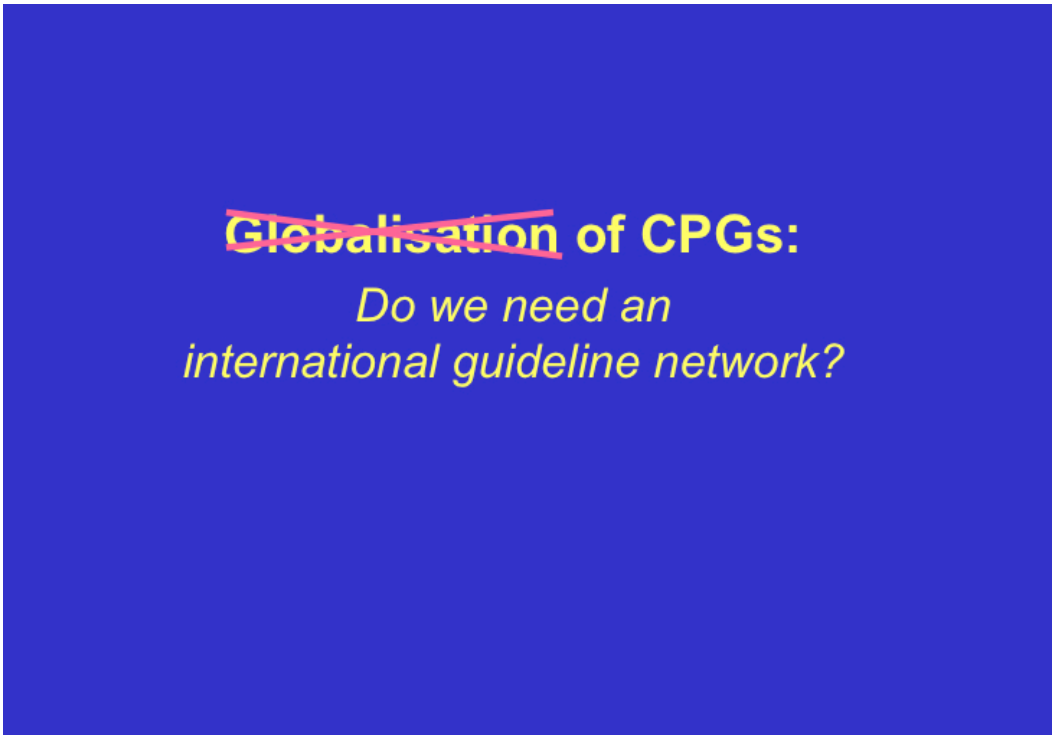


Figure 3. J Miller & G. Ollenschläger, “Globalisation of CPGs: Do we need an international guidelines network?” (slide 17) Presentation in Berlin, June 2002. Available at: www.aezq.de/mdb/edocs/pdf/azq-veranstaltungen/cpg/miller.pdf

In November 2002 the Guidelines International Network is officially registered as a not-for profit “international network of excellence for the research and implementation of guidelines” (Ollenschläger et al., 2004: 456), with 22 founding committee members from 12 countries. Since then GIN’s membership has grown to “93 organisations and 89 individual members representing 46 countries from all continents”, and its website has put guideline development quite literally ‘on the map’, by providing hyperlinks to all its members. (www.g-i-n.net/about-g-i-n, accessed 09-02-2012).

GIN’s founding documents explicitly state an opposition to the standardization of guidelines, and instead see diversity as a resource for “building partnerships” and “to promote information sharing and cooperation

in guideline development" (Guidelines International Network, 2002). GIN promotes its methodological standards (AGREE, GRADE, ADAPTE) as 'ideals' for all guideline developers to live up to, but they are voluntary standards: GIN accepts members regardless of whether they comply with GIN's standards, and no sanctions or enforcement are in place. The development of formal certification has been discussed at GIN meetings, but remains controversial. Jako Burgers, founding member and former chair of GIN opposes certification and sees the role of the network to "connect organizations like a kind of highway, a kind of infrastructure, and in that way tries to improve communication and exchange." (interview 24-08-2010, Chicago). This echoes EU's policy to establish networks of excellence as *social connections* between people to advance research. This model considers diversity a resource for innovation, and is uncoordinated with the broader policy of harmonization, which is associated with an alternate model of networks as *shared infrastructure of technical standards* aimed at facilitating harmonization (Barry, 2001: 89-93). Despite GIN's opposition to standardization as a formal and explicit policy, its standards and knowledge practices do standardize guideline development.

GIN standardizes guideline development

The Guidelines International Network creates voluntary standards and comparative knowledge with which guideline developers self-regulate their practices through standardization. GIN standardizes based on the same normative pressures of emulating what is considered best practice as described above. GIN's annual conference provides an important forum for guideline developers to measure, compare and rank their own and other's guideline development practices along methodological Gold Standards such as AGREE. But it is not only the exchange of knowledge that *results* in standardization. Standardization is also *required* to establish knowledge exchange in the first place. A session at GIN's 2011 annual conference

illustrates this. The speaker compared several approaches to grading evidence, without finding any of them superior. But instead of concluding that diversity in grading systems was legitimate and harmless, attendees preferred having a uniform approach in order to: understand and learn from one another; interpret and judge each other's evidence summaries; reduce efforts by using each others' evidence summaries (fieldnotes, GIN conference 2011, Oral Session 3.10). Since variation itself is considered an obstacle to improve (and reduce) guideline developers' work, attendees preferred the GRADE system simply because many guideline developers around the world already use it. Since standardization is required for establishing knowledge and collaboration, GIN's goals of "improved communication and exchange" result in standardization. Such shared standards may not represent "the cheapest, most efficient, safest, scientifically most reliable, or technically most advanced outcome" but may be established by repeating "how things are already done by most parties" (Timmermans & Epstein, 2010: 79).

Adherence to methodological standards may be cumbersome, time consuming and expensive, but GIN's growing membership suggests members expect such constraints to be overshadowed by its benefits. As Timmermans and Berg have shown, shared standards can be appealing because they increase the legitimacy of criticized or unrecognized work. For weak professional groups "the process of standardization forms an attractive strategy to rally members and claim expertise." (2003: 93). Standardization is particularly attractive when standards are developed and monitored from *within* the professional group, and they display enough flexibility to allow for professional judgment and discretion. GIN aims to transform guideline developers from a heterogeneous, unorganized group whose work is criticized by medical professionals (who denounce standardization) and policy makers (who demand more standardization), to an emergent profession that claims its own domain of expertise. Following AGREE helps

defend guidelines against criticism (and globalization) because the process of arriving at the guideline is 'up to' standard. According to the subtitle of its logo, AGREE advances "the science of guideline development", allowing guideline developers to lay claim to *scientific* procedures. GIN makes experts in guideline methodology visible as a distinct group of 'guideline methodologists' and provides a platform to learn and display their skills, expertise and procedures. The standards do not provide universal rules that can be mechanically followed, and judgment and discretion of guideline developers remains necessary. And since there are no fixed norms or sanctions, standards may be endorsed as ideals without (fully) living up to them. Failure to do so may be used to petition guideline financers for more resources to bring guidelines 'up to standard'. Indeed, the extensive (and expensive) methodological requirements advocated for explicitly encourage guideline production to take place within centrally organized and well-funded organizations that are fulltime in the 'guideline business' (Grol et al., 1998: 265). The standards exclude local groups of medical professionals – disapprovingly referred to as 'GOBSAT' or Good Old Boys Sat Around a Table' (Guyatt, interview 7 October, Hamilton) – to develop guidelines without 'professional' guideline methodological experts (Hirsch & Guyatt, 2009). By encouraging the standardization of 'professional' and 'systematic' guideline development, GIN's methodological standards thus result in a rejection of both global and local guidelines and a strengthening of the development of national guidelines.

Conclusion

By developing quality standards for guidelines and the establishment of the Guidelines International Network, EU funding created a new governance structure where the voluntary nature of its standards did not limit its regulatory success, but made it successful. By connecting humans,

practices, technologies, standards and knowledge across disciplinary and national borders this socio-technical network is neither ‘bottom-up’ nor ‘top-down’ standardization, but is a coordination of social, organizational and technical elements across local, national and international levels. GIN’s regulatory impact is distributed along those elements: both more ‘localised’ than EU regulatory structures by effecting the daily practices of individual guideline methodologists, and also operating far beyond the EU by guiding guideline development in South Korea, North America and the WHO. Yet, despite initial hope and promise of universal Gold Standards for health care, this new governance structure did not result in increased Europeanization of health care. Not because it lacked the “power” to regulate and standardize across Europe, but by changing *what* was being regulated and standardized. The vision of guidelines that regulate physicians directly with ‘universal evidence’ is rejected, insisting “norms and values” co-determine guidelines. By measuring the quality of guidelines by its methodological process, guidelines do not require uniformity of clinical content and the rationalization of guidelines is disconnected from the standardization of guidelines. Experts in guideline methodology are posited as a new (self-regulated) profession indispensable for ensuring the “objectivity” of medical regulation. This new governance structure can be considered a new type of meta-regulation in medicine, distinct from medical professionals self-regulation and from governmental legislative powers (Cambrosio et al., 2006). It pursues meta-regulation by aiming for improvement in the quality of quality improvement instruments, but rejects meta-standardization by legitimizing variation in guidelines. This meta-regulation of guidelines further entrenches guideline development at the national level because it rejects global guidelines, and its extensive quality standards discourage local guideline development.
Since GIN does not consider clinical practice guidelines to be universal objects, it rejects the establishment of European guidelines as a ‘technical’ infrastructure to standardize health care practices. Instead, GIN is supported

by, and supportive of, EU policies that aim to establish 'social' networks in which diversity is considered a resource for learning and knowledge production. However, collaboration amongst diverse stakeholders both produces and requires shared infrastructure and objects, therefore the notion of sociotechnical network is more productive to capture these dynamics than are models of technical or social networks. The case of GIN illustrates how a sociotechnical network can transform local practices around the world with voluntary standards. Yet, without an EU institution (or other central actor) determining who will enroll in such networks, and how standards are going to be used (if at all), the EU has little control over the kind of standardization that a network will produce. Since standardization is not always the emulation of a Gold Standard, but may be established by the lowest common denominator or a consolidation of existing practices, there is no guarantee that standardization will result in rationalization or Europeanization. If the EU pursues the establishment of networks as a policy goal in itself, with little attempt to coordinate harmonization policy with networks, European New Governance may be highly effective, but the EU will have little control over *what* it is effective at.

CHAPTER 6

Patient & Public Involvement and evidence based guidelines: four conceptual models

Abstract

Patient and Public Involvement (PPI) initiatives are increasingly put into practice across all types of health care settings, from policy making to medical research, from hospital administration to clinical activities. By affecting what clinicians offer as 'standard practice', guidelines influence both the care of individuals and the (re)design of health care systems, and PPI in guideline development holds the promise of achieving care that is both evidence-based and patient-centered. There exists both enthusiasm and skepticism about what PPI can contribute to Evidence Based guideline development, but many guideline developers agree that PPI has become unavoidable and seek out guidance about the purpose, role and methods of PPI in guideline development. Based on literature on PPI in health care and medical science, supplemented by empirical material from participant observation, interviews and document analysis of those involved in the promotion, planning and practice of PPI in guideline development, this article presents four conceptual models of PPI in guideline development. The focus of these conceptual repertoires is the rationale for PPI. It details promises as well as limitations of each model, clarifying what the purpose of participation is, who needs to be included, and what it is they can contribute. The four rationales are, first, Consumer Choice, which draws on consumers' right to choose their own and personalized care. Second, Democratic Voice, which draws on the rights of citizens and taxpayers to base policy on public values. Third, Lay Expertise, in which patients' experiential knowledge of care, context and illness contributes to 'context-sensitive' guidelines. Fourth,

Critical Witness, in which patients provide epistemological oversight to strengthen the Evidence Based approach, ensuring accountability to evidence. Finally, I argue that diversity in PPI is not an obstacle to establishing PPI as a 'universal standard', but an essential feature to achieve PPI.

Introduction

Evidence based Medicine (EBM) is at the center of contemporary reforms to make medicine more scientific, providing clinicians with quantitative, 'universal' evidence of best practice, aiming to replace (unwarranted) practice variation with 'gold standard' treatment (Timmermans & Berg, 2003). EBM has been criticized for being "doctor-centered" and "erasing the patient" (Mykhalovskiy & Weir, 2004), its emphasis on evidence from population studies leading to a standardization in which the needs, preferences, characteristics and experiences unique to individual patients are ignored (Rogers, 2002; Greer et al., 2002; Van der Weijden et al., 2010). EBM is considered to be competing and incompatible with the strong reform movement of 'patient centered care' (PCC) (Armstrong, 2007; Bensing, 2000). The latter aim to make medicine more humane by encouraging patients to make their own decisions about their care, and participate in decision making in "priority-setting, the planning of services (including resource allocation), and the delivery of services" (Wait & Nolte, 2006: 153-4). By changing what clinicians offer as 'standard practice', Evidence-Based guidelines reform both the care of individuals and the design of health care systems. By involving patient and public representatives in guideline development and use, holds the promise of achieving care that is both evidence-based and patient-centered.

Organizations that develop evidence based guidelines have been encouraged to include patient representatives in guideline panels from the

first landmark report on guidelines by the Institute of Medicine (Field & Lohr, 1990) to its recently updated standard for guideline development (Institute of Medicine, 2011). The National Institute for Health and Clinical Effectiveness (NICE) [13] in the UK has an extensive program of PPI in place, which has functioned as a model for other guideline developers around the world. In 2007, about 60% of guideline development organizations surveyed had a policy for patient involvement in place (Topalian & Pagliari, 2007). And other guideline developers are getting ready to take that "next step in guideline development" (Krahn & Naglie, 2008: 436). In 2010, at the annual conference of the Guidelines International Network (GIN) almost 400 guideline developers from around the world debated the promises and pitfalls of PPI. The conference was announced as a "breakthrough for PPI", as it was concluded that the main question is no longer *whether* to do PPI, but *how* to do it, as both enthusiasts and skeptics agreed that PPI is unavoidable (Fieldnotes, 28-08-2010).

This article proposes four conceptual models that describe the promises and limitations of patient and public involvement in guideline development (Table 2). A diverse literature exists on patient and public involvement in guideline development, including articles advocating for the importance of PPI (Kelson, 2001; Boivin et al., 2010); (grey) literature providing practical advice on a range of PPI methods (CBO & NPCF, 2009) and overviews of PPI in practice (Van de Bovenkamp, Grit & Bal, 2008). Like literature on PPI in general, this literature is marked by "a lack of conceptual clarity", its abstract, vague and contradictory nature being an important barrier to putting patient and public involvement into practice (Church et al., 2002: 13). Some of the articles focus narrowly on a very specific type of PPI, such as single PPI initiatives (Van Wersch & Eccles, 2001), the inclusion of 'preferences' (Boivin et al., 2009; Chong et al., 2009); or the

[13] www.nice.org.uk/getinvolved/patientsandpublic/patientandpublichome.jsp

involvement of patients in their own clinical care (Van der Weijden et al., 2010), while other articles mention a great variety of benefits, reasons and methods of patient and public involvement, but lack conceptual clarity about the who, how, why and to what end of PPI. The two studies that have reviewed much of this literature focused on evidence of effectiveness without distinguishing diversity in the *purpose* of participation, *who* is included, or *what* they contribute (Van de Bovenkamp & Trappenburg 2009; Légaré et al., 2011). So while a rich literature exists on patient and public involvement on guideline development, it contributes to, rather than alleviates, the "conceptual muddle" of PPI models (Forbat, Hubbard & Kearney, 2009: 2547). To provide more clarity, I propose four distinct models, or 'repertoires' of PPI in guideline development (Table 2).

	Consumer Choice	**Democratic Voice**	**Lay Expertise**	**Critical Witness**
Guidelines	Objective information for patients (patient decision aids)	Policy documents, normative instruments	Knowledge tools for professionals	Any of the previous
Who	Autonomous consumer	Disinterested citizen	Professionalized patient	Critical skeptic
Contribution	Individual preferences	Collective values	Experiential knowledge	Epistemological oversight
Purpose	Free choice Personalized care	Democratic policy	More useful guideline	More objective guideline
Limitation	Standardization of preferences Marketization of care Options pre-determined	Contradictory values Lack of interest Threat to guideline validity	Lack of representative lay-experts	Maintenance of independent, critical view Reactive

Table 2: Summary of the four models of Patient & Public Involvement in guideline development

These four models are based on a literature review of PPI in guideline development, supplemented by existing models of PPI in other health care settings, and illustrated by empirical material of the way guideline developers understand, practice and object to PPI. The focus of these conceptual repertoires is the *rationale* for PPI: clarifying what the *purpose* of participation is, *who* needs to be included, and *what* is it they contribute. While practical examples are given, the four rationales are mostly conceptual models of the logic of or reason for PPI. As such, this chapter primarily aims to contribute to existing literature by supplementing models

for PPI in individual care and policy-making with rationales for lay involvement in science (drawn from literature in Science & Technology Studies), and by suggesting a new model. An additional aim of this chapter is to facilitate PPI in practice. I hope to provide guidance to guideline developers by increasing their understanding of distinct rationales, addressing objections to and limitations of PPI models, and highlighting the benefits of diversity in models and methods.

The first and second models reiterate the conceptual 'consumerist' and 'democratic' models of existing literature on patient and public involvement in health care. These draw respectively on consumers rights (access to full information and freedom to choose) in health care decisions on an individual level and the 'social' rights of citizens (and taxpayers) to democratic decision-making on a collective or policy level (Wait & Nolte, 2006; Florin & Dixon, 2004; Tambuyzer et al., 2011; Tritter, 2009). These rationales apply to PPI in guideline development when guidelines are either considered to be decision-aids for the use of patients or are consdered documents used to reform and redesign health care services (Veenendaal et al., 2004; Van de Bovenkamp & Trappenburg, 2009), and can be reconfigured in various ways to advocate the inclusion of patient *preferences* in guidelines (Boivin et al., 2009). In addition to these classical ideals of descriptive-statistical and democratic representation, 'expertise' provides a third rationale for patient and public involvement (Martin, 2008). Expertise provides the basis for particpation in guideline development when guideline development is considered 'evidence synthesis' for the education or strategic benefit of medical professionals. Drawing on the literature on 'lay' involvement in science, this model can provide a 'technocratic' rather than democratic rationale for 'lay experts' to contribute unique experiential knowledge to guidelines. The final model proposes a fourth rationale for 'lay' participants in Evidence-based knowledge activities. Modified from what Lehoux, Daudelin and Abelson call a 'reflexive witness' (2012), the 'critical

witness' performs a distinct regulatory role, provding epistemological oversight to ensure accountability to evidence and the objectivity of experts.

After mapping these four repertoires and illustrating them by literature, interviews and practices of PPI in guideline development, I address the implications of the diversity of these models and argue that diversity in PPI repertoires is *not* a barrier to patient and public involvement. While diversity may hamper (standardized) evaluation, it is necessary for successful PPI.

Material & methods

The four rationales for PPI are based on literature review of PPI in medicine and science in general, as well as on (grey) literature on PPI in guideline development specifically. This review did not seek to be comprehensive, but rather to capture the range of rationales of PPI in health and medical settings that have been described in the literature. The literature review is supplemented by empirical material gathered from interviews and observations as part of a larger study conducted in 2007-2010 examining methods of guideline development to illustrate how the four rationales are drawn upon in the practice of guideline development. Relevant empirical material from this study consists of fieldnotes taken as participant observer at five subsequent annual conferences of the Guidelines International Network (GIN), a meeting attended by around 400 guideline developers and researchers from 85 organizations from around the world. This included two workshops (2009 and 2010), three plenary lectures, and dozens of oral sessions specifically dedicated to PPI. From 2008-2010 I acted as participant observer at GIN's working group on Patient and Public Involvement (G-I-N PUBLIC) and once data collection was completed in 2010, I was elected Co-Chair. I held semi-structured interviews which lasted about an hour with five members and staff of patient organizations involved

with guideline development (denoted as *PtRep1-4*), including the person responsible for guideline development activities at *Nederlandse Patienten Consumenten Federatie* (NPCF – Dutch Patients Consumer Federation), the largest federation of consumer and patient organizations in the Netherlands (*PtRep5*). I interviewed staff responsible for Patient & Public Involvement activities at the Dutch Institute for Healthcare Improvement (CBO), one of the largest developers of multidisciplinary guidelines in the Netherlands, and active developer of guideline methodology, both nationally and internationally (denoted *GuiDev1-2*). Finally, I interviewed the program manager for the public funding for guideline development in the Netherlands (ZonMW), which has mandated PPI as a requirement to receive funding (denoted *Funder*). Interviews were recorded and transcribed in full. I observed meetings of six different guideline development groups (GDGs) at the CBO in the Netherlands. Three of these groups (GDG1, GDG2, GDG6) included one or more patient representatives. Their meetings were audio-recorded and, based on fieldnotes, specific passages of interest were transcribed. The notations used to report this data refer to the guideline group (*GDG1-6*) and the number of meeting (a,b,c). I translated material that was in Dutch into English, and to maintain confidentiality of participants certain information in direct quotes taken from GDG meetings was altered (such as the guidelines' topic or kinds of treatments).

Consumer's choice: informed and free choice in personalized care

The 'consumers choice' discourse draws on consumer rights (access to information and free choice) and is concerned with fully informing patients so they can make autonomous decisions, and with ensuring the medical market provides services that fulfill the needs and preferences of health care consumers. The (expected) benefit of this patient involvement is choice,

autonomy and better care as it is tailored to patients' needs. Moreover, it aims to regulate and improve (and facilitate) the health care market, as only services that consumers need and prefer will thrive (Wait & Nolte, 2006: 152-53; Tritter, 2009).

Within this discourse, guidelines can been considered a threat to individualized consumer's choice as their strong prescriptive language makes treatment choices pre-determined by a guideline panel instead of by doctor-patient dyad in the clinic (Barratt, 2008; Boivin, Légaré & Gagnon, 2008), and the standardization of medical practice ignores legitimate treatment variation due to patients' diversity in preferences and needs (Greer et al., 2002). There is concern Evidence Based Medicine will dictate physicians' behavior and reduce patient choice, guidelines prescribing 'Gold Standard' treatment on the basis of quantitative evidence of narrowly defined patient groups, outcomes and treatments, recommending "what was best for the overall group rather than for the individual" (Rogers, 2002; Murphy, 2008: 232). Guidelines are to be reformed from standardization tools regulating the behavior of physicians to decision aids that provide standardized information on therapeutic options, not to be 'followed' or 'complied with' by doctors, but that they and their patients discuss, interpret, modify or ignore, making decisions shared, evidence-informed and value-based (Boivin et al., 2009; van der Weijden et al., 2010). Two kinds of personalization can be distinguished. First, to achieve "professional care" (Boivin et al., 2009), standard advice is to be tailored to clinical characteristics such as "age, sex, disease severity, overall risk profile, and combinations of comorbidity" (Krahn & Naglie, 2008: 436). Secondly, to achieve "informed choice" guideline recommendations need to be adjusted to the patient's life situation and preferences, for example preference for treatment process (weekly visits to a clinic or daily medication) (Boivin et al., 2009). To ensure patients are involved in personalizing standards in the clinic, additional tools such as patient versions and patient decisions aids

are being developed (Raats, van Veenendaal, Versluijs & Burgers, 2008), and patient organizations producing their own guidelines primarily aimed to inform patients and improve shared decision-making (PtRep4, interview 27-08-2010), aiming for guidelines to "no longer be a 'forcing, demanding' instrument" but "primarily a tool to start a dialogue with the patient" (personal communication, PtRep1).

The consumer choice model can also be framed as *aggregated* consumer choices that form the 'demand' side of the health care market. In a privatized health care market responsibility for the quality and price of care is transferred from government regulation to the self-regulation forces of the market where well-informed consumers' will 'vote with their feet' by choosing the care (and carers) that best respond to their needs. This is a common discourse where governments pursue increased privatization and marketization of health care services such as the Netherlands and the UK (Hughes et al. 2009; Mol, 2006). Guidelines are envisioned to improve the functioning of the healthcare market by better aligning healthcare 'supply' and consumers' 'demand' (van Veenendaal, 2004: 51). Moreover, alongside other evaluation projects (e.g. performance indicators, patient satisfaction surveys), guidelines can increase democracy by providing citizens with quantified and evaluative information previously only accessible to experts with specialized expertise (Porter, 1995). By providing consumers with an evaluative yardstick to judge the value or quality of health care (Van de Bovenkamp, Grit & Bal, 2008: 14,43), guidelines become one of the "calculative devices for reconfiguring marketized healthcare as value-driven" rather than "cost-saving" (Zuiderent-Jerak, 2009: 780).

This discourse was strongly expressed by the employee of the *Nederlandse Patienten Consumenten Federatie* (NPCF – Dutch Patients Consumer Federation) responsible for work on patient involvement in guideline development, and the preferred organization to work with in many guideline development projects. In 2006, as part of a new health care

insurance law aimed at increasing marketization in health care, the Dutch government funded a range of projects to be undertaken by the NPCF with the explicit purpose to strengthen the position of patients as the "third party" in the health care market, becoming equal partners vis-à-vis providers ("who own the knowledge"), and private insurers ("who have the funding"). Guidelines can 'empower' consumers because guidelines "lead to [performance] indicators. And those indicators allow you to see transparency and quality." The performance data enable consumers' choice by comparing providers' performance on the 'Choose better' website launched by the government ("www.kiesbeter.nl"), which also provides 'decision aid tools' to help patients make specific treatment decisions. Moreover, patients are encouraged to use (patient versions of) guidelines as a yardstick to evaluate their own care, "Finally you know how [care] *should* be. […] patients can bring it into their doctor's office and ask "Why does my care not follow the guideline?"" (PtRep5, interview 12-06-2008). Guidelines may provide a basis on which to discuss, direct and challenge the care patients receive and to ensure patients know how to choose 'gold standard' treatment

Both versions of this discourse are largely concerned with producing (modified) guidelines to allow patients to be involved in their own personalized care. Some suggest patients' preferences should be included during the *production* of standards, and several methods to do so are proposed. The participation of a consumer representative in guideline development groups is sometimes suggested, but in line with the EBM evidence hierarchy, rather than the 'opinions' of a single consumer representative, the inclusion of 'evidence' on patient preferences is often suggested as a more reliable method to know of patient preferences. Various methods may collect such evidence: surveying a statistically representative sample of patients for preferences; qualitative research (focus groups, interviews) to establish patient preferences; conducting a literature search for existing research on preferences (Chong et al., 2009;

Murphy, 2008; Greer et al., 2002; Díaz del Campo et al., 2011). Others question how aggregated data on preferences can be useful to inform standardized recommendations, as a patient's personal preferences should inform their own care "regardless of the existing population-based preference data" (Umscheid, 2009: 988). A guideline developer expressed concern that finding '62% of patients prefer A' would lead a guideline panel to recommend A as the only option, while many patients do not in fact prefer that option. He objects to such use of "standardized" preferences, and rather involves consumer representatives on guideline development panels not to provide 'average', 'majority' orr 'standard' preferences, but to ensure diversity is maintained in the guideline, and hence in clinical practice:

> "if you want patients to participate in a guideline group you want people who can manage differences. Not those who are 'average', nor those who advocate one opinion, but those who know there are divergent opinions and know how to work with that. You can always distinguish some categories, but the task of the patient representatives is precisely that some things should *not* be standardized, instead you have to ask the patient." (GuiDev2, interview 12-06-2008)

When patient representatives were asked to evaluate the options the guideline panel could recommend, they suggested the best recommendation would be to "discuss with the patient, meaning, offer a patient a choice between the options" (GDG1, comment from patient representative on draft guideline, p.3). Other participants in the GDG can also play this role. For example, when GDG2 debated whether GPs or specialists would make the best "care coordinator", a nurse suggested the guideline development panel recommends that individual *patients* choose their own care coordinator (GDG2d).

Limitations

Limitations of equating patient involvement with ensuring consumer choice are well-rehearsed in literature on PPI in health care. PPI is criticized for allowing public institutions and physicians to withdraw from ensuring the quality, efficiency and affordability of care and transferring a collective responsibility to individual patients or the local community (Mol, 2006; Tritter, 2009: 279). It is limited to encouraging choice between pre-determined options, and allowing the public to evaluate the output of services, but does not encourage patient and public involvement in determining the options or quality of the services (Tambuyzer et al., 2011: 7). For patient and public involvement in guidelines in order to affect the content, kind, number, quality and accessibility of the options that will be available in the clinic in the first place (Bastian, 1996), we turn to the 'democratic perspective'.

Democratic voice: basing policy on public values

In the democratic voice model the rationale for involvement draws on the democratic rights of the public to participate in policy making (Wait & Nolte, 2006:152; Forbat, Hubbard & Kearney, 2009: 2549; Tritter, 2009: 281-2). This rationale is particularly powerful in publicly-funded initiatives where members of the public are not only (potential) users and recipients of services, but also the financial stakeholders (as taxpayers). PPI can be requried in order to receive public funding to develop guidelines, as is the case for NICE in the UK, and ZonMW in the Netherlands. In the case of privately funded guideline development, the participation of public representatives can provide a mechanism to produce a decision-making tool that better reflects the 'values' of the public as recipients of care. The production of a decision-making tool always involves technical as well as normative questions (Boivin, Légaré & Lehoux, 2008), such as what risks are worth taking and how to balance competing goals and interests. Moreover, by changing standards of care, guidelines can change not just the

care of the specific patients in question, but also effect the design, delivery and resource allocation of other health care services. Within this rationale, guidelines are considered normative and regulatory tools, requiring to complement the 'technical' evidence (provided by experts) with the values of (potential) patients and the general public.

The "values of patients and society" are to provide a more legitimate basis for the normative aspects of guidelines, because the 'values' of experts are affected by financial and professional conflicts of interest (Boivin & Légaré, 2007: 1308), "the public are the experts on values, not physicians or policy-makers." (fieldnotes, GIN conference plenary 3, 24-08-2007). Due to their personal and financial interests in receiving the treatment of their choice, patients (and their repesentatives) are often considered unable to provide input on 'collective values', particularly decisions about resource allocation[14]. When considering the *relative* value of health care interventions across many conditions, it is the public (voters, taxpayers, citizens) that are to provide input. To distinguish them from the 'special interests' in a specific condition that patients and professionals share, 'ordinary citizens' are largely defined by what they are not: disinterested and dispassionate, not activists, patients, researchers or health professionals (Lehoux, Daudelin & Abelson, 2012: 1844). To represent the values of 'the public' at large they are expected to represent (and be recruited from) a wide range of geographic, demographic, ethnic and socioeconomic groups affected by the policy (Martin, 2008). As part of the extensive program of PPI set up at the National Institute for Clinical Excellence and Health (NICE) in the UK, a Citizen Council "a panel of 30 members of the public that largely reflect the demographic characteristics of the UK" exists that is not involved in the

[14] One could distinguish values from preferences by distinguishing a 'universal' principle (i.e. patient autonomy) from a specific expression of that principle (i.e. let patients choose their care coordinator). Patients' "values" can then often be classified as "preferences" that fall under the previous model (e.g. Do patients value/prefer a tiny survival benefit despite large toxicity of the treatment?).

development of specific guidelines, but provides advice on the “social value judgements” that should guide the development of all guidelines[15]. The Council is asked to formulate general principles, such as “principles that should govern the imposition of public health measures on the UK population” and “What should NICE take into account when making decisions about clinical need?”. Other than NICE, few guideline development organizations have organized involvement of such ‘disinterested’ citizenry.

Limitations

There are a host of limitations associated with this repertoire. They include problems to reconcile or combine values of “patients *and* society” that are in conflict with one another; and the difficulty of interesting ‘disinterested’ citizens to be involved, especially public representatives outside a self-selected subgroup of white, middle-class, highly educated individuals (Martin, 2008), as well as the validity and legitimacy of their perspectives. Questions are also raised about the impact of this kind of PPI on the quality of guidelines. While the inclusion of public representatives may improve the quality of the decision-making process (more fair and democratic), this does not guarantee a good (valid) decision. The quality (truth, clinical validity) of a guideline (and the resulting health care recommendation) may suffer when ‘values‘ are being prioritized over ‘evidence’. A ‘democratic’ development process may become the primary end in itself, regardless of the effects on the quality of a guideline (Van de Bovenkamp & Trappenburg, 2009: 205). Such an ideological or ‘principle-based’ stance is expressed by the public funding organization that mandates PPI :

> “I think in the Netherlands we agree that patients have a seat at the table, one way or another. Regardless of whether it has an impact or

[15] http://www.nice.org.uk/aboutnice/howwework/citizenscouncil/citizens_council.jsp

not. [...] To me it doesn't matter, it can simply be a belief you have. Patients have to participate, it's about them after all. It's about them, they receive the care, they have to have a voice. And I don't think the guideline is going to be worse. Well, who knows what [evaluative] research will find." [..] but at least then we know. Either we can say 'it is a real contribution, the guideline improves.' Or we simply agree together that 'the care process is multidisciplinary and the patients are one of the parties at the table. Period.'" *Funder, interview 19 –10-2009*).

A more conceptual limitation of this model is the demarcation of technical knowledge (evidence) provided by professional experts from "values and preferences" provided by patient/public, which neglects that patients and the public can also contribute knowledge and obscures that professionals also bring subjectivity and normative assumptions to the guideline development process (Lehoux et al., 2009:2005). These latter limitations are addressed respectively by the 'lay expert' and 'critical witness' model.

Lay expertise: experiential knowledge for contextualized guidelines

The previous rationales for PPI aimed for guidelines to achieve either more personalized care or more democratic policy. But guidelines may be produced to be neither patient decision-aids nor policy documents, and may rather consider medical professionals as their only audience. Medical specialty associations can produce guidelines as 'evidence synthesis' for the education of professional members (Armstrong, 2007), or as strategic tools to re-organize intra-professional jurisdictions and responsibilities (Timmermans & Berg, 2003; Castel, 2009). Neither consumers' right to choose, nor the democratic right of citizens to be heard provide convincing

rationales to participate in knowledge tools produced by and for medical professionals. This third model of emphasizes patient's knowledge and expertise to justify their involvement even in the most 'technical' aspects of medicine, such as the design and conduct of research, or the training and education of medical professionals (Martin, 2008). In this model, 'values' are not considered a distinct category to be added to the knowledge of experts, but central to constituting what are 'technical issues' and scientific knowledge in the first place (Jasanoff, 2003). In contrast to the former model, the demarcation of the 'technical' from the 'political', deciding what the salient issues are, determining the range of solutions to be explored, and the kinds of knowledge considered necessary to answer the question are at once political and scientific. These issues should not be determined exclusively by experts, but established through discussion with the public in which the distinctions between facts and values can be challenged and the boundaries between experts and the public blurred (Wynne, 2003; Lehoux et al., 2009). Although the lively debates around Evidence Based Medicine, and its Evidence hierarchy, have mostly taken place between professionals (philosophers, social scientists, medical professionals, epidemiologists) without extensive public controversy or debate, it illustrates the political nature and consequences of how what counts as 'Evidence' is defined, and how it directs the questions and solutions explored (Lambert, 2006; Mykhalovskiy & Weir, 2004; Goldenberg 2006).

Epstein describes how AIDS/HIV activists were successful in acting as 'lay experts', gaining access to discuss the technical, epistemological and political issues with professionals as equal partners in dialogue. In order to interact and converse as credible partners with experts they familiarized themselves with biomedical language and methods (Epstein, 1995). This does not mean lay experts are not critical of the professional collective they seek to become part of. While not directly opposing or rejecting biomedicine, they do seek to transform it. Patients' experiences with the disease and

health care allows them to assert their own set of questions, concerns and knowledge, “re-opening the list of issues that should be taken into consideration in medical practices beyond those brought in by medical science” (Akrich et al., 2012: 32). Their unique experiences as patients provide them with “knowledge about their own bodies and its pain; the ways in which the body reacts to medication; and intimate knowledge about the circumstances in which they live.” (Prior, 2003: 53). Such knowledge can provide insight into problems, harms or benefits that typically go unmeasured in randomized clinical trials. This includes the ‘work’ patients (or their carers) have to do (traveling to the hospital every week), and 'invisible' symptoms such as changes in fatigue, appetite, libido, fear, hope or pain which can greatly affect the patient’s family, work, emotional, social or sex life. In GDG2b, in absence of a reliable diagnostic test, the best way to measure the cognitive decline of patients was to “ask the patient’s partner about it”. And in GDG1, the patient organization that had reviewed the guideline draft suggested that monitoring of the condition should be more intensive than “what is objectively necessary” (GDG1, patient representative’s commentary of guideline draft). Although not ‘medically’ necessary, more intensive monitoring would help reduce patients’ fear of not being treated adequately and in time, a fear which itself causes serious psychological, social and emotional problems. CBO acknowledges the need for knowledge about the patients’ daily lives that goes beyond ‘the medical’, by stating that medical professionals may evaluate the quality of a medical treatment, but it is patients who can best evaluate the quality of life (Raats, 2009: slide 13).

Limitations

The contradictory status of being simultaneously lay and expert provides challenges for this model, leading to concern that the professionalization required to interact with (and as) experts may lead to “non-representation of

the patient community and to the loss of the 'pure' experiential knowledge" (Caron-Flinterman, Broerse & Bunders, 2005: 2582; Van de Bovenkamp, Grit & Bal, 2008: 18,42). Two kinds of 'non-representation' can be distinguished, one that questions the validity of experiential knowledge when patient representatives disagree with professional's experiences, and another that questions the ability to defend patient's interests when patient representatives agree with professionals.

Questioning the validity of lay experts' expertise

Professionals may argue that the experiential knowledge of professionalized patients is based on *selective* experiences, not representative of, shared by or valid for 'average' patients that are not active within or member of a professionalized patient organization (Van de Bovenkamp, Grit & Bal, 2008: 35-6), and professionals' may represent their own impressions of the experiences of patients they see daily in their clinic. Patient representatives attempt to give their claims weight by demonstrating they are not personal experiences or concerns, but are shared by many others, as shown in the following exchange during a GDG meeting:

> *GDGmember:* it's difficult with experiences to-
> *PtRep:* oh but these are not personal experiences, this is a problem that has been identified by the [patient] association. (GDG2a).

Nonetheless, the formal EBM procedures that provide rules on how to assess the quality and validity of different kinds of knowledge, places all experiences at the bottom of the Evidence hierarchy, casting doubts on the credibility and legitimacy of patients' knowledge and expertise, demanding additional rationales (statistical or democratic representativeness of model one or two) to earn a seat at the table. Moreover, experiential knowledge is always secondary to evidence (RCTs, large series, etc.) Both these 'lessons' can be challenged, which does not require guideline developers to

abandon the evidence hierarchy, but requires an understanding of methods of producing 'experiential knowledge', and the placement of experiential knowledge – like other 'non-evidence' - outside the hierarchy all together.

First, whether experiential knowledge is valid for (or 'representative' of) patients' experiences depends primarily on whether experiences have been collected, aggregated and formalized from a broad and diverse patient constituency (Rabeharisoa & Callon, 2004). If member surveys have been done, they can speak for (and about) the experiences and opinions of a large number of patients, including those who are less educated or professionalized or hold divergent ideologies (Akrich et al., 2012: 14-17). Individual patient representatives can not claim descriptive-statistical or electoral representation, but their personal professionalism or education does not confine their contribution to "the opinion of a small, articulate group" (Bovenkamp & Trappenburg, 2009: 209). Having knowledge of a broad range of experiences, they can claim 'experiential representation' (Martin, 2008: 37). The solution to reliable experiential knowledge and representation is not to recruit 'average' patients without knowledge of medical language and scientific methods, limited organizational and communication skills, and limited contact with other patients. All of these qualities are needed to formalize experiences into knowledge, and be able to 'represent' those experiences during the expert discussions within GDG meetings. Valid experiential knowledge does not require "pure" patients, but "patient representatives should stay in close contact with the patient population they represent, verifying the mutuality of demands, ideas, and judgements regularly." (Caron-Flinterman, Broerse & Bunders, 2005: 2582). A medical professional may (claim to) possess all these qualities, but lacks access to patients' concerns and experiences beyond the clinical encounter, which is what makes experiential knowledge unique in the first place.

Second, 'lay' knowledge, based on lived experiences, is situated and contextualized, rather than 'universal' and abstract (Wynne, 1992). Instead of weakening its credibility, this quality can provide a unique and necessary role to mediate between the 'universal' evidence of RCT results and the 'local' and situated nature of decision-making in 'real world' practice. The 'transfer' of facts or artifacts from research to clinic is never a simple question of 'applying' results to practice but requires a process of translation and re-contextualization (Mykhalovskiy, 2003). As chapter four showed, diverse knowledge and justifications, what CBO calls 'other considerations', are needed to assess whether RCT evidence – valid under very specific experimental conditions - is relevant, applicable and valid in the 'local' circumstances of the clinical practice guideline. To select, apply and contextualize evidence of trial results, guideline developers rely on a range of knowledge outside the Evidence Hierarchy: ethical and biological principles, knowledge of the current standard of practice, knowledge of the specific clinical and other characteristics of the patients under consideration (Knaapen, in press). Such 'non-evidentiary knowledge' is not subordinate to evidence but plays a mediating role between 'universal' evidence and 'local' recommendation. And chapter five emphasized that, in line with calls to "globalize the evidence, localize the decision" (Eisenberg, 2002), 'local' guidelines are becoming the gold standard of guideline development (Knaapen, 2012). Guidelines' translation of 'universals' to 'local' settings also aims to make them more used and useful in practice, because for standards to be workable at all, they need to be incorporated into existing (net)works, personal and institutional routines and concerns (Timmermans & Berg, 2003). The production of such "context-sensitive guidance [that] shows both what works and how it might be implemented in specific circumstances" is associated with seeking "the expertise, views, and realities of stakeholders" (Lehoux et al., 2009: 2004) so that guidelines can take into account the views and practices of its intended users. Patient

representatives can claim stakeholder status as 'guideline users', claiming a seat at the table not based on representation, nor on the lived experience of illness, but on their unique knowledge of the *context* of the health care where guidelines are to be used. Like other stakeholders and guideline users (policy makers, nurses, GPs, insurance physicians) they provide promise to improve the use of guidelines that are better adjusted to the realities, needs, knowledge and beliefs of those who are to follow the guidelines' recommendations (PPI workshop 25-08-2010; PtRep5; PtRep4, interview 27-08-2010). Guideline developers promote patient's knowledge as a means to produce guidelines that are more relevant and useable in practice, as reflected in CBO's definition of patient participation:

> "contributing specific experiential knowledge by patients or representatives of patient organizations, the purpose of which is to achieve equal influence on the development and execution of a project/ activity and to improve acceptance and implementation of the products" (CBO, 2007: K6-7).

Questioning lay experts' interests

Patient advocates' concern about the representativeness of professionalized patients is different. They worry that working extensively with professionals leads 'lay-experts' to adopt the professional's concerns, no longer able to be critical of them (Rabeharisoa, 2006), creating a schism between the concerns of 'lay-experts' and 'lay-lay' patients (Epstein, 1995). A consumer representative for the Cochrane Collaboration (that produces systematic evidence reviews) writes about the difficulties of remaining critical of the 'colleagues' you are closely cooperating with. She emphasizes this is not just related to funding, but also due to "the direct experience of inclusion--being given a seat at the table—[…] I know from firsthand experience that it's all too tempting to become cooperative and 'play nice.' It's not easy to remain an activist, challenging the conventional wisdom, when you are

treated as a colleague--or given that impression, anyway." (CCnet online discussion, dd.09-05-2009). Patient advocates in the Netherlands raised such concern about the cooperative stance of the Federation of Patients and Consumer Organisations in the Netherlands (Nederlandse Patienten Consumenten Federatie, NPCF), the umbrella organization that guideline development and governmental organizations work with almost exclusively (see Akrich et al., 2008:184 for 'preferred partnerships' with large federations). Patient advocates not part of NPCF criticized NPCF for simply following the terms and conditions set by government and guideline developers (PtRep1, interview 13-02-2008; PtRep4), reflecting 'polderization', a Dutch term for finding consensus through collaboration that poses the danger of a loss of strength and activism (Akrich et al., 2008: 182). The NPCF staff member explains how she was unable (and unwilling) to incorporate its members' comments into a handbook on patient involvement in guideline development that NPCF had co-written with CBO, as they had already determined the handbook's purpose and tone. NPCF considered their members' demands (for more funding and *real* involvement) suitable for a "political manifesto", but not a guideline development manual (PtRep5).

To 'represent' patients' rather than professional 'interests' it is important lay experts retain an amount of independence (financially, and otherwise) from other professional parties (doctors, guideline developers, public and private funders) to allow them to define and design their priorities, activities and input independent from their professional partners (Van de Bovenkamp, Grit & Bal, 2008: 46). The role taken on by the NPCF is not primarily to identify and 'defend' patient interests against those of professionals or government, and has no ambition (or authority) to take a leading role in guideline development (GuiDev2, PtRep5). NPCF's role is more akin to the mediator role described by Rabeharisoa, facilitating the emergence of new and shared interests, alliances, knowledge and entities

between various social actors (funders, medical professionals, guideline developers, patients/research subjects) (Rabeharisoa, 2006). Under the constraints and encouragement of public policy and funding, patient organizations, funding agencies and guideline development organizations cooperated to each perform their role in the health care market under construction, and NPCF's primary vision was the creation of 'consumer choice' as a regulating market force.

While the two distinct concerns about the representativeness of 'lay experts' are often raised, neither implies that professional skills deny, erase or oppose experiential knowledge *in principle*. The real difficulty with the role of the lay expert is that in practice few patients or their organizations have the resources to be able (and/or willing) to fulfill all three conditions: experiential representation, relevant professional skills and independence. The difficulty of patients to participate as lay-experts, has led some researchers to reconsider the patient's presence in GDGs entirely (Van de Bovenkamp & Trappenburg, 2009), but some patient representatives do participate actively as 'lay-experts' (GDG1, GDG2), even if their contributions are not marked as such in the final guideline, which is a collectively written text, as chapter three made clear (Knaapen et al., 2010). However, in GDG6, the patient representative (as well as another 'stakeholder', a pharmacist, on that group), made very few contributions, and the two suggestions she made were considered 'off-topic'. She felt she lacked the medical and statistical know-how to act 'like an expert' within the GDG, and would have preferred to be consulted in a separate meeting. Alternatively, Collins and Evans suggest that "experience-based experts" who lack interactional competence can contribute by finding a mediator (a social scientist or an advocate) to translate their experiences in a manner that the expert-experts understand and can take into account (2002: 256). To help ensure lay experts are able to contribute on behalf of many patients, patient organizations aim to recruit members who demonstrate professional

skills (scientific background, public speaking or committee work) and who can draw from other patients' experiences through meeting with or conducting research amongst the other members of a patient organization (PtRep5). Interviewees emphasized reliance on a broad range of patient experiences by having multiple patient representatives on a guideline committee and holding focus groups or (online) consultation with other patients, as literature has recommended (CBO & NPCF, 2009; Van Wersch & Eccles, 2001). Finally, even when patient representatives infrequently contribute to GDG discussions, the fourth "critical witness" model provides reasons why they may still have an important impact on both the process and end-product.

Critical witness: epistemological oversight

The fourth model provides a final rationale for PPI in "Evidence Based" activities, proposing a kind of public oversight or regulation of the 'values' and knowledge that expert GDG participants bring to the table. Lehoux, Daudelin and Abelson's recent empirical work (2012) has suggested citizens involved in a science/policy network in genetics acted as a 'reflective witness' (turned voyeur), ensuring transparency of the network's deliberations. Several interviewees and workshop participants also saw merit in patients witnessing the GDG process, even without active participation: "as a layperson you are amongst a group of medical specialists. They normally always see you from one side of the table, as a patient. Suddenly you are one of them. […] it gives it a 'face': "there's one!". (PtRep3, interview 10-06-2008). Such presence makes the language of the debate and final guideline more 'patient-centered', for example using the term 'people' instead of patients or sufferers (GDG2d). It may also improve intra-professional consensus, either because professionals are uncomfortable disagreeing in front of patients, or by ensuring that a clear

and common language is used instead of jargon that is unclear to experts and care providers in other specialty areas (PPI workshop 25-08-2010; PtRep4).

Being a 'mere' witness does run the risk of providing legitimacy to decisions predetermined by 'expert' participants, without impact other than rewording the recommendation in more palatable language. The role of a *critical* witness is to question and check the legitimacy of decisions, by demanding justifications from experts. This demands accountability of experts that is not so much 'political' as epistemological. By "insisting on accountability to evidence" (Brouwers, Somerfield & Browman, 2008: 1026), critical witnesses can provide a type of regulation or 'public oversight' that aims not so much for transparency or democracy, but for safeguarding the evidence-based approach and objectivity of experts. As chapters four and five demonstrated, ensuring the 'EBM-ness' of guidelines does not mean purging it from all 'subjective' judgments and values, but providing procedural rules to manage evidence, non-evidentiary justifications and the selection and conduct of experts (Knaapen, in press; Knaapen 2012). Since experts' specialized knowledge is widely recognized to be indispensable, but 'tainted' by financial and professional stakes in the topic, "biases cannot be eliminated but rather should be examined openly and pruned of unreasonableness." (Cohen, 2004: 418). Thus, "nonexperts who may have less knowledge but fewer of the influences that contribute to bias" are called upon to ensure 'objectivity' of guidelines (Detsky, 2006: 1033), including independent methodologists (Hirsch & Guyatt, 2009) and chairs as "reasonably partial spectators" (Cohen, 2004). Patient and public representatives may also play this role, since:

> "In fact, consumer representatives are often strong advocates of an "evidence-based" approach. Faced with competing views or monolithic unchallenged assumptions and interpretations, it is often a consumer representative who will say "prove it"." (Bastian, 1996: 488).

This role avoids the need for statistically representative consumer preferences, democratically representative public, and the 'experiential representation' of lay experts. It does not aim to exclude values, preferences or (financial) interests from guideline development (as in model one), nor to replace professionals' values with *the* values of patients and society (as in model two), but expects 'lay' participants to critically question the assumptions and values made by professionals for making recommendations (Schünemann et al., 2006: 2of8). The role of a critical witness is distinct form those of experts, and does not require experiential knowledge or medical training of the 'lay expert' model. To demand accountability to evidence, the critical witness does require understanding of EBM principles, and maintain a critical stance and skepticism in order to question participants, rather than having blind trust in experts, evidence or EBM rules. The guideline developers or patient advocates in my study did not explicitly express such an epistemic regulatory role for patient or public participants, but Akrich and colleagues describe the involvement of a French childbirth organization in guideline development as an evidence-based activism that questions the evidence-base of clinical practices and challenges the 'political' nature of certain guideline recommendations (2012: 20). Consumers of the Cochrane Collaborative Network also expressed the role and qualities of consumers in systematic evidence reviews in such terms:

> "I think intelligent skepticism and respect for evidence is the essence of Cochrane and an informed skepticism, *not* personal experience, is in my mind the most important thing consumers can bring to the table." (emphasis in original)

> "Patients may not know medicine, but they know how to smell a rat."
>
> Contributors to Cochrane Consumer Network email list

And while this role avoids some of the limitations of the other models, the challenge to remain skeptical and critical of the professionals and evidence that the participant works so closely with, is even more critical for the critical witness than it is for the lay expert. Others might object that this role is rather limited in its impact, as the oversight role is reactive, rather than a proactive contribution (Tritter, 2009).

Diversity or standardization of PPI

The diversity in PPI models raises the question which of these methods is the best one. In 2007, the Guidelines International Network (GIN) established a working group on patient & public involvement in guideline development (called G-I-N PUBLIC). In line with GIN's mission to establish international methodological guideline development standards (Knaapen, 2012), GIN recommended "the development of standard measures to assess the effectiveness of patient and public involvement policies" in order to "develop standards on how best to involve patients and the public in guideline development and implementation." (Boivin & Marshall, 2008). Establishing the best PPI method requires evaluation, and evaluation requires standardization. Standardization of PPI thus seems essential to establish highest quality PPI.

A recent attempt to measure PPI's impact found no empirical evidence that PPI enhances the quality of guidelines (Van de Bovenkamp & Trappenburg, 2009). Yet, an "absence of evidence should not be mistaken for an absence of effect" (Crawford, 2002:1266), as only very few documents *reported* any impact of PPI on the guideline, its development or implementation or on health outcomes (Légaré et al., 2011), a difficulty encountered by other studies examining the effectiveness of PPI in health care and medical research (Nilson et al., 2006; Crawford et al., 2002). The

difficulty in evaluating whether (or which) PPI is effective is compounded by the diversity of PPI rationales, as there is no agreement on *what* constitutes successful involvement (Lehoux et al., 2009; Wait & Nolte, 2006). Since each of the four PPI models has a legitimate purpose (i.e. personalized care; democratic policy; better guideline quality; more objective guideline), presents both promises and pitfalls, prioritization requires evaluation. Moreover, it is unclear how to define or measure each of these outcomes in a standard manner (e.g. chapter five discusses the difficulty of determining the quality of a guideline), and even less clear how to measure the impact of the patient or public representative on the outcome. Chapter three demonstrated how the final text of a guideline is the result of a collective process of debate that is framed by procedural rules and the reflexive critical contributions of all GDG members (see also Moreira, 2005). Evaluation thus faces important challenges that can not be remedied by standardization, and it may well be impossible to determine, not just in practice, but in principle, how a collective emergent guideline process is affected in a predictable way (either positive or negative) by a single participant, be they lay or professional.

Standardization of methods is also supported to facilitate the practice rather than the evaluation, of PPI. It is argued that if GDG participants shared a single definition and purpose of PPI, and agree on the characteristics, role and contribution of 'lay' participants *before* embarking on PPI activities, disagreement and conflicts around PPI will be reduced (Veenendaal et al., 2004; Boivin et al., 2009). Standardization of methods may indeed be an effective strategy for successful collaboration amongst actors with divergent practices and 'world views' (Star & Griesemer, 1989), however, standardization itself can be so difficult, that instead of collaboration, communication and understanding, it leads to disagreement, conflicts and contestation over what the single standard should be (Timmermans & Epstein, 2010). Standardizing of PPI methods requires

much more than the creation of standard terminology of PPI (consumer, patient, public or citizen; involvement, participation or engagement). Agreeing on a single best PPI method implies (and requires) agreement on a host of other concepts, ideals and practices: what are guidelines, what is knowledge/evidence, what is (good) medicine. Actors from across Evidence Based Medicine and Patient Centered Care could come to agree on terminology of PPI, they still diverge a great deal on the role and meaning of (good) medicine, knowledge and guidelines. Since each of the four PPI models presents a legitimate purpose (i.e. personalized care; democratic policy; better guideline quality; more objective guideline), and its own promises and pitfalls, prioritization would require evaluation.

Thus, standardization is not only difficult to achieve, it is no panacea to either evaluation or collaboration. Collaboration between diverse actors may instead be achieved by the development of what Star and Griesemer (1989) call 'boundary objects'. Boundary objects "have different meanings in different social worlds but their structure is common enough to more than one world to make them recognizable, a means of translation" (Star & Griesemer, 1989: 393). Since a boundary object has flexible meanings, divergent actors can establish a 'mutual modus operandi' that does not require the resolution of all disagreements and prioritization of a single shared 'worldview'. Instead of an obstacle to bringing diverse participants together, diversity in meaning can be a virtue to bring different communities of practice together around one concept. Epstein also found that looseness in terminology and flexibility of practices can function as a *strength* when building coalitions across diverse groups (Epstein, 2007: 88). In terms of PPI, the conceptual clarity of the four models can provide a shared structure to PPI, but practice can accommodate various ways of doing PPI, and agreeing on a single model is not necessary. Just as participants in GDGs can hold diverse concepts about the purpose of guidelines and draw on a range of repertoires of evaluation (Moreira, 2005), GDG participants can

hold divergent notions of PPI. Chapter three illustrated how pragmatic solutions were found without solving conceptual differences, moving the production of a guideline forward while agreeing to disagree (Knaapen et al., 2010). Disagreements can be solved by informal and pragmatic debates, not pre-determined position-taking. And opportunities to find pragmatic, local agreement and cooperation increase when PPI can mean *many* things. For example, CBO prefers to use the term "patient perspectives" because it may indicate patient's values, experience and/or knowledge (GuiDev1). 'Policing' PPI practices to ensure participants' characteristics match their appointed role and contribution (and correcting or excluding them when it does not?) is both impractical and unnecessary because diverse roles overlap and distinctions are ambiguous. Successful practices of PPI do allow for 'unexpected' contributions; someone recruited with the characteristics of a 'disinterested citizen' can uphold consumers' right to choice and access (see Will, 2009: 624). PPI will be more effectively and more easily managed when a range of roles, participants, contributions, identities, concept of guidelines are permitted.

Since its establishment, G-I-N PUBLIC has shifted from the idea of an 'international standard' to the metaphor of a methodological "toolbox" (G-I-N PUBLIC Steering Committee, 2012). For guideline developers who are convinced (or legally mandated) that PPI is the 'right thing to do', the Toolbox is to provide methodological advice on "doing it right" (Boivin, 2012: 12). But unlike an international standard, a toolbox implies that various PPI methodologies can be applied, tried out and experimented with depending on local context (fieldnotes, G-I-N PUBLIC workshop, 2010). Instead of finding agreement on what 'good' PPI 'really' is, this international hybrid forum of patients, public, guideline developers and researchers now considers variation in PPI methods as legitimate. This shift is in part the result of the lack of proof (according to EBM's evaluative standards) of effectiveness of any single PPI method (Van de Bovenkamp & Trappenburg,

2009; Légaré et al., 2011). But the lack of 'universal' success of PPI methods has also served to reinforce the idea that the success and quality of PPI methods intrinsically depend on 'local context' :

> The choice of a specific method of involving patients and members of the public in guidelines is not only a technical decision on "what works best", but also reflects socio-political choices regarding the organisation and governance of health care delivery. Sensitivity to local context should guide any work that aims at harmonising methods of involvement at the international level." (Marshall & Boivin, 2008: 7)

G-I-N PUBLIC thus promotes not just *guidelines* that are 'context-sensitive' and adjusted to local needs, but also promotes localized and 'context-sensitive' guideline development *methodology*. This contrasts with the Guidelines International Network's main vision, described in the previous chapter, that supports guidelines that are 'context-sensitive' and 'local', but stresses the *methodology* to develop such guidelines is *universally valid* (Knaapen, 2012; Qaseem et al., 2012).

While a lack of standardized PPI methods will be an obstacle to the *measurement* of the impact of PPI, maintaining vagueness and diversity may facilitate successful *performance* of PPI. Paradoxically, for PPI to be effective, we need to stop focusing on measuring PPI's effectiveness that demand PPI (and guideline development) to be standardized from place to place and over time (Learmonth, Martin & Warwick, 2009: 106). For PPI to be performed successfully in guideline development - which is a contextualized and collective achievement – it needs to be 'local', diverse and flexible. Preservation of the "rich complexity" of PPI is important to make patient and public participation worthwhile (Lehoux et al., 2012: 1849). By providing conceptual clarity based on four conceptual models, aims to provide a structure which will avoid PPI being *so* flexible that 'anything

goes', leading to tokenism where *any* kind of PPI is simply and automatically good. Diversity in the methods and practice of PPI may be the best recipe for success, but the cooks must be provided with the right tools and know their ingredients well.

CHAPTER 7 Conclusion

Standards that avoid standardization

This dissertation has presented a detailed empirical analysis of the production of evidence-based guidelines, including meta-standards that regulate guideline development. Each chapter has presented unique, yet overlapping findings, which differ from much existing literature by emphasizing the pragmatic nature of evidence, and the importance of meta-regulation provided by 'guidelines for guideline development'. In concluding, I will summarize the main findings and address how they fit within existing debates in the literature. First, I address how the findings recast the nature of evidence according EBM principles, and the changing nature of the latter. Secondly, I address the emergence of meta-regulation as a distinct kind of regulation, different from of other types of regulatory control in medicine. Meta-regulatory instruments (and meta-regulators) provide a novel way to ensure the quality control of quality control instruments. And thirdly, I will provide a synthesis of how this answers the main research question I started out with; how do guideline producers justify and adapt their work in response to the multitude of promises and critiques that surround EBM guideline development? It is the development of 'universal' standards for 'local' guideline development that is crucial in managing tensions between universal and local, between evidence and values, between standards and individuality, between science and care. Meta-regulation of duplicate guidelines challenges the polarized notions of universal gold standards as well as dehumanized global uniformity. It is also distinct from other notions that refer to the management of the gap between 'universal' evidence and the complexity of clinical practice, both those further 'downstream' (the use of standards) and further 'upstream' (the production of evidence), such as local universality, situated standardization and niche standardization.

The pragmatic nature of evidence and EBM

This study has shown that what counts as evidence in the production of a primary EBM tool is not determined solely by the formal and 'decontextualized' evidence hierarchy. Unlike the formal hierarchy, the classification of evidence in practice does include assessment of external validity, or relevance. A variety of 'non-evidentiary' considerations (biological, ethical, pragmatic, local) have to be relied upon to determine what counts as evidence in the first place. Morever, more than (pragmatic) evidence is needed to formulate the guidelines' recommendations, including non-evidence, procedural rules, and an alignment with the 'external world' such as external review by experts and practitioners and the incorporation of existing clinical standards.

The existing literature on evidence according to EBM has missed these "important and desirable pragmatic features of the evidence-based approach", not as a result of focusing primarily on what counts as evidence (as Goldenberg, 2009: 168, suggests), but because they ignore how categories of (non-)evidence are constructed in practice, instead equating what counts as evidence with its 'textbook account' portrayed in the formal evidence hierarchy. Work by the late Leigh Star showed how formal classifications have important social, political and philosophical consequences, yet it also highlighted that the erasure of pragmatism and situatedness is an inherent feature of any 'universal' category, as Latimer aptly summarizes: "classificatory systems themselves can be understood to hide more than they reveal: they are reductions that efface the complexity and messiness of medical categorizing" (Latimer et al., 2006: 604). The categorization of actual cases - whether the diagnostic classification of patients or categorizing studies as evidence - is never simply the 'application' of formal criteria, but is a practical accomplishment that produces emergent categories (Latimer et al, 2006; Bourret & Rabeharisoa, 2008; Hedgecoe, 2002). It is therefore perhaps unsurprisingly that my empirical investigation of the sociotechnical practices that allow for the categorization of evidence finds a more complex, pragmatic and situated account

of what counts as evidence (and non-evidence). I do not argue that this discrepancy between pragmatic practice and formal rules implies that guideline developers' support for (or claim to) EBM is mere 'rhetoric', or that their practices ignore EBM principles. This study contributes original insights by demonstrating how guideline developers manage the local contingency and pragmatism *within* the constraints set by EBM principles.

Guideline developers do not deny or hide the pragmatic and situated nature of their practices, nor consider it a flaw to be remedied by more universal evidence, or by stricter (following of) EBM rules. Rather, contextualized pragmatism is embraced and promoted as good practice. A host of guidelines for guideline development are developed that aim to reconcile EBM's formal principles with the pragmatism and situated judgment of informal practices. Such meta-standards modify both formal rules and informal practices, as they act as what has been called regulative rules (regulating existing practice) as well as constitutive rules (creating new practices) (Cambrosio et al., 2009: 657). Distinct meta-standards include the principle of evidence searched guidelines that define EBM as the elimination of intentional ignorance. This redefines the nature of EBM from the presence of evidence, to a distinct management of the absence of evidence. The GRADE procedure seeks to replace traditional evidence hierarchies (limited to methodological design) with a more contextualized classification and assessment of evidence. This includes a formal assessment of relevance, allowing for down- and upgrading of evidence based on various 'judgments'. And thanks to a 'situated intervention' (Zuiderent & Jensen, 2007) by a STS researcher, the lack of procedural standards for 'non-Evidentiary' considerations was addressed at the most recent annual meeting of the Guidelines International Network, inciting "wide interest in a dedicated working group for further developing methods for weighing and including different types of knowledge in guidelines." (Zuiderent-Jerak, Forland & MacBeth, 2012: 2of2). Thus, new procedures are emerging to integrate different forms of evidence into CPGs, developed and supported by practicing guideline developers who are vocal proponents of EBM, including the 'founding father' of

EBM himself. It must be emphasized that it is the *procedures* that are new here, not the pragmatism in evidence classification or the diversity of knowledge relied upon in guideline development. It may thus be too quick to confirm suggestions that a "new EBM" is developing that is "more open to the integration of different forms of evidence" (Goldenberg, Borgerson & Bluhm, 2009: 165). The legitimacy of more pragmatic evidence is achieved by claiming more transparency. Yet, this visibility does not simply 'reveal' informal practices unaltered, but requires formalization and standardization of what was informal and locally specific. Particularly, an empirical study of the GRADE procedure in use will be important to understand how existing guideline development practices will be transformed, as 'transparency' may obscure more than it reveals.

Regulation of guideline development

The second issue this study addresses is the question of 'quality control of quality control tools' (Power, 2000). Many accounts exist that outline the diverse, and contradictory, regulatory impact of guidelines. Evidence based guidelines have been claimed to provide regulatory control or professional autonomy to almost any group in medicine: young doctors can use them to challenge senior clinicians; academic researchers can use them to dominate 'rank and file' clinicians; managers or government can use them to control medical professionals; medical professionals can use them to prevent control by third parties, or competition from other professional groups; industry can use them to promote pharmaceuticals over non-medical solutions; patients can use them to keep their doctors 'up to standard'. The 'true' power of EBM and its guidelines resides in this *variety* of possible regulatory effects. Thus, every professional and regulatory group in the health care domain aspires to the production of their own EBM tools and standards. The resultant proliferation of standards leads to the need for a regulation and quality control of duplicate and contradictory standards. This dissertation demonstrates

how the Guidelines International Network (GIN) emerges specifically devoted to the quality control of these quality control tools.

GIN does not standardize the 'double standards' by developing specific universal gold standards. Instead, it legitimizes EBM guidelines as a class, by developing meta-standards that regulate the standard-setting process, while maintaining diversity in (national) guidelines. Instead of defining the quality of EBM guidelines in terms of Evidence (as academic 'knowledge' elite might claim), or on a democratic basis (as policy-makers might claim), meta-standards measure a guideline's quality by the 'transparent reporting' of its production process. The focus on transparency as a requirement *demanded* of guideline production, instead of a result *produced* by guidelines, is a new point of departure from which to analyze EBM amongst increasingly important "technologies for transparency" (Blomgren & Sahlin, 2007: 161). The transparency demanded by meta-standards assumes that guideline users monitor guideline development, and hold guideline developers accountable for the 'objectivity' of their evidence selection and expert judgments. Putting the quality control of guidelines (and the 'objectivity' of those developing them) in the hands of the physicians who are expected to submit to those same guidelines, suggests the endless regression of 'control over control' (Power, 2000) is avoided by establishing a type of co-regulation between users and producers of guidelines.

The meta-regulation thus provided by EBM confirms that regulatory relations in guideline development are not necessarily "along traditional professional lines (i.e., nurses versus doctors); instead, it could be that support for clinical practice guidelines generates heterogeneous alliances of health care actors." (Timmermans & Kolker, 2004:189). Not only does meta-regulation differ from other efforts to control medical practice (i.e. state, managerial, or professional regulation) (Cambrosio et al., 2006), a new heterogeneous alliance of health care professionals has emereged to produce and promote meta-standards. The meta-standards developed by the new international guideline network has brought together experts from very diverse professional groups (librarians, epidemiologists,

general practitioners, psychologists), declaring themselves a unique and united group of experts in guideline development methodology, or 'guideline methodologists'. Instruments such as AGREE and GRADE function as international standards that legitimize and professionalize the work of this heterogeneous group of guideline methodologists, in a similar fashion that standards have strengthened the professional status of 'weaker' professional groups such as insurance physicians or nurses (Timmermans & Berg 2003; Berg et al., 2000: 773-5). Since the meta-standards are based on the existing practices of GIN's founding members, they mostly reflect and explicate existing work practices rather than modify them. And the standards are ambiguous enough to leave enough interpretative space that requires the discretion and understanding of experts in guideline methodology: following the meta-standards requires expertise and judgment. Meta-regulation thus allows guideline methodologists to claim unique professional expertise and territory[16] in guideline development methodology, an independent role that – in addition to clinicians, medical researchers, and patient representatives – is indispensable to ensure high quality, 'objective' and 'scientific' guidelines. Two decades ago, the first prominent guideline 'methodologists' noted that "the *art* of developing practice guidelines is in an early stage" (Field & Lohr, 1990: 57, my emphasis), by now AGREE, GIN and related instruments claim to have developed a veritable international *science* of guideline development. Meta-regulation has shifted regulatory relations within medicine by specifying the roles of 'academic elites', 'stakeholders', and guideline users in guideline production, and by making 'guideline methodologists' – not academic elite, clinicians, hospital managers or government agencies - the experts on quality control of quality control instruments.

---

[16] This 'profession' primarily exists on an international level, no national guideline development associations or education exist. It would be an interesting issue for further research to compare if, in this globalized world, other professions have first emerged in an international arena, after which they continue to establish national educational degrees, professional associations, annual meetings, etc. Emergent professions might follow a reverse course of the international connections of many national professional organizations.

Gold Standards that avoid standardization

The notion of meta-regulation is key to answering the overall question of my dissertation: how do guideline producers justify and adapt their work in response to the multitude of promises and critiques that surround EBM guideline development? Meta-regulation emerges when the proliferation of guidelines threatens the notion of universal gold standards, and international collaborations respond by celebrating variation, treating the heterogeneity of guidelines as a necessary virtue "to better meet divergent patient and physician demands in a caleidoscopic world." (Pentheroudakis et al., 2008: 2077). Instead of universal evidence or reproducible clinical content, they measure the quality of guidelines in terms of adherence to procedural standards, invoking transparency and standardized procedures as the cornerstones of high quality evidence based guidelines. One of the elements of the development procedures is the involvement of patients and the public, which aims to 'bring back' the patients, values, choice and preferences that standards threaten to erase. By developing these 'universal' meta-standards, uncontrolled guideline production is regulated without the pursuit of uniform medical practices. It is these specifics of meta-regulation that allow for rationalization of guidelines without standardization of medicine. Quality control and accountability (of guideline development) is achieved while standardization (of guidelines, judgments and clinical practice) is avoided. Meta-regulation simultaneously invokes the imperative to "valorize local rule and respect difference" and the claim that "standardized processes are often more transparent in ways that are consistent with accountability." (Timmermans & Epstein, 2010: 82). By promoting national guidelines, both sides of the polarized debate on evidence based guidelines are challenged, as 'local' gold standards simultaneously pursue and limit standardization. Meta-regulation does not rely on traditional scientific virtues of 'mechanical' objectivity that expunges all 'subjectivity' to produce reproducible outcomes, but the meta-standards allow for diversity in guidelines and pragmatism in their development process. 'Local' gold standards aim to avoid both

the uniformity of a globalized medicine devoid of humanity, as well as the positivist utopia of universal gold standards based on 'context-free' evidence that require the exclusion of cultural beliefs, local routines and human judgment.

Despite the legitimacy of diversity and the need for 'judgment' in their development, 'local' guidelines are still to function as gold standards that regulate *illegitimate* practice variation and clinical judgment. The new EBM procedures of guideline development help define, measure and report which variations and judgments are legitimate, and which ones are not. As such, the universal standard-setting processes reflect "regulatory objectivity" which "includes decisions about which components of a particular configuration of action should or should not be standardized, as well as the means and extent of such standardization." (Cambrosio et al., 2009: 655). For example, the systematic evidence search defines what absence of evidence is legitimate, moreover, its reporting procedures serve to show how that legitimacy was established. The 'judgment' that is formalized and legitimized in EBM guideline development is not the judgment, opinion or experience of individual practitioners, but the collective judgment of guideline developers familiar with – and disciplined by - specific evidence tools (i.e. GRADE, AGREE). Critical judgment is required of individual practitioners (as guideline users), as they are expected to assess the legitimacy of the guidelines (and its transparency reports). Finally, *what* practice variation is legitimate, and which ones are to be replaced by the standard, continues to be debated by guideline developers. For example, in one guideline group there was debate whether variations in treatment between patients in rural and urban areas was illegitimate practice variation, and the guideline was to ensure rural patients also got access to the 'standard' treatment. Or was this to be considered 'context' to take into account, and would a different treatment (for which rural patients do not need to travel for 2 days) be a better treatment standard for rural patients?

The latter example illustrates that the notions of 'local gold standards' and 'contextualization' of evidence provide no answers as to what the relevant 'locality' or 'context' is, be it continental, national, regional, institutional or individual (the

latter would suggest ‘personalized’ standards such as patient-decision aids). The aim of ‘local’ gold standards is not to achieve compatibility of several different clinical practices, for example in different regions or countries. Rather, guidelines are expected to make treatment of patients more standard or equal *within a* clinical practice, and compatibility is sought between different specialties, hospitals, funding arrangements, etc. To this end, guidelines aim to integrate different entities (other guidelines, existing routines, legal and financial arrangements) within the context of clinical practice. While meta-regulation rejects globalization of CPGs in name of diversity in “organizational and cultural contexts”, national guidelines are expected to integrate and connect different organizational elements (guidelines, protocols, regulations, standards and categories) into a single guideline. The Guidelines International Network emphasizes the legitimacy of variation between countries, by assuming that ‘one’ clinical practice (should) exist on a national level (as well as an even more problematic assumption that culture and values are national), thus both assuming and promoting homogeneity within a nation. At GIN’s annual meetings, each individual is treated as a ‘representative’ of a country, displayed on name tags[17], and reinforced by moderators in workshops who ask “How do you solve problem X in Australia?”. Notwithstanding multinational research practices, globalization of ‘culture’ in general, and much talk of medical tourism and global health, individual patients do rarely cross national boundaries for their treatment (and if they do, transitions and compatibility is not (expected to be) smooth). Yet, the notion that clinical practice is organized and regulated on a national level is as much an assumption about reality as a goal to be achieved by developing and promoting national clinical practice guidelines. Not only do other organizations gladly pursue global guidelines in order to regulate and coordinate national practices on an international level (e.g. WHO guidelines on pandemics and 'global health' issues) or to globalize clinical practices (e.g. the Global

[17] My nametag would usually display ‘Canada’, even though as a Dutch national living and working in Quebec while researching guideline development in France, it was highly problematic which nation I represented.

Guidelines Task Force of the 'World Gastroenterology Organization' (Fried et al., 2006)). Even within GIN, a working group called G-I-N North America was set up in 2011, to better accommodate the specific 'context' of clinical practice and guideline development in North America (or rather, the USA[18]). Compared to the centrally organized and publicly funded health care in many European countries, the organization and regulation of clinical practices in the USA is more fragmented and relies more on market regulation and private organizations. Nationally developed guidelines are promoted to achieve the centralized regulation and coordination that is lacking in this system (Shaneyfelt, 2012). Currently, while national coordination may be lacking, the more relevant 'context' or 'system' within which a patient's care is highly organized in the USA seems to be within hospital/insurance networks, that already provide highly integrated clinical care practices. For example, Kaiser Permanente is an American GIN member that promises "seamless" and "coordinated care" by providing diagnostic labs, prevention, care, referrals, insurance, electronic health records, and trial research all within a single system[19]. Such institutional systems are internally highly integrated, but compatibility between networks may be absent, since their interrelations are characterized by competition rather than cooperation, differentiation may be more important than compatibility. By promoting and regulating nationally developed guidelines, meta-regulation not just reflects, but also pursues the organization and regulation of clinical practices on a national level.

Managing the tensions between binaries

Meta-regulation, by legitimatizing local standards and formalizing 'contextualization' of evidence, can be considered a novel answer to the question whether 'context-free' evidence or 'universal' standards are valid in 'local' settings.

---

[18] Canadian health care arrangements resemble the 'organizational contexts' in many European countries in that they are centrally regulated and publicly funded, (primarily) by provincial governments.

[19] http://thrive.kaiserpermanente.org/experience-better-care (accessed December 10, 2012).

This question is not only important to the epistemological politics of EBM guidelines, but a central and contested issue in almost all standardization processes. Other scholars have described related, but distinct, ways to bridge the contentious gaps between the experimental and controlled environment of RCTs (and their selective research subjects) and the complexity of clinical practice (and their diverse and unique patients). These related processes occur both further 'downstream' (in the use of standards), and further 'upstream' (in the production of evidence) from guideline production.

Work on the *use* of standards has demonstrated that standardization of practices is always incomplete, for standards to be workable within the 'disorder' of existing (net)works in practice, informal work-arounds and local adjustments to the standards are necessary (Hogle, 1995; Timmermans & Berg, 2003). Thus, standards do not simply standardize practices, practices also localize a standard, in other words only 'local universality' can be achieved (Timmermans & Berg, 1997). My findings show localized gold standards are promoted, and demonstrate that during their development guidelines incorporate the standards, materialities and expertise of existing clinical practice. Thus, the notion of local universality is not only the result of using standards in practice, but is the goal that guideline developers pursue from the start. We could say guideline developers pursue local rationality, as they support the idea that best practices are locally defined.

Expanding on the notion of local universality, Zuiderent-Jerak has suggested a process of 'situated standardization' (2007). Since standardization is more effective when standards are not developed in isolation of the practices in which they will be used, Zuiderent-Jerak suggests a process of 'situated standardization'. This means the distinct phases of first designing/producing standards and then implementing them in practice is abandoned, and change is directly produced from situated interventions and experimentations in practice (May (2006) uses the term "practice based evidence" to denote reliance on local experimentation and qualitative evidence). Situated standardization thus avoids an endless proliferation of 'implementation tools' of evidence synthesis, guideline recommendations,

clinical pathways, regional care standards, institutional work procedures and patient-decision tools. The author claims these tools frequently fail because they simply repeat and privilege the aggregated knowledge (both normative and clinical) of the 'universal evidence' over the knowledge and norms of practice, thereby not reducing the gap between science and practice, but reproducing it between the new tool and practice (Zuiderent-Jerak, 2007). The practices of meta-regulation of local standards described in this dissertation differ from 'situated standardization' described by Zuiderent-Jerak. This is not only because the 'locality' of standards differs (national versus institutional), since the determination of 'local' is not fixed by meta-regulation. What differs is that meta-regulation does not abandon the distinction between standard-setting and standard-use, as it aims for universal standard-setting processes, not situated standardization through local experiments.

The often-repeated mantra "globalize the evidence, localize the decision" illustrates my claims. While universality of clinical practice is not the objective, a distinction between evidence (considered universal) and the *use,* application or implementation of evidence (considered local and diverse) is maintained. Yet, guidelines are neither global evidence nor localized decisions, as Busch aptly puts it, standards are used to connect language to the material world (2011: 3). Guideline developers' intentions (and practices) are to connect categories of universal evidence and clinical practice, not to abandon them altogether. Thus, guideline developers' discourses maintain the constructed binaries universal and local, evidence and practice, but do shift and blur their borders. They create hybrid practices, tools and new categories that connect one category to another. The notion of universally developed localized gold standards defies binary classification, and is a hybrid tool specifically intended to connect 'science' and 'care'. Thus, while the notions of 'universal evidence' and 'local practice' are maintained, the practice of guideline development shows that evidence is not treated as 'universally valid', or as 'context-free', but as dependent on the very specific (and controlled) context in which it was developed. I would argue the many 'implementation tools' currently created along the ever-longer chain of 'knowledge

translation' don't simply 'recreate' "the very concerns they were supposed to address." (Zuiderent-Jerak, 2007: 312), because they do not simply transfer 'globalized evidence' unaltered along the chain of implementation tools, but a transformation takes place. Each tool provides a different process of disentanglement and qualification of evidence (Moreira, 2007), and the existing standards get "swamped" with the disorder of practice as they are adjusted to and aligned with the routines, materialities, expertise and needs of care practices (New & Moen, 2010). Rather than simply repeating the implementation gap, the gap gets transformed as it moves along the chain of textual translations from 'universal' evidence reviews (Moreira, 2007), to national guideline recommendations, to regional care standards, to 'integrated' clinical pathways (Allen, 2009), to institutional work procedures (Moen & Nes, 2012) and 'personalized' patient-decision tools. In order to link knowledge production to the regulation of medical practices, each implementation tool aligns 'universal' evidence with various complexities of care practices, and/or connects several tools and practices with one another, creating an interconnected network of 'nested' standards that becomes difficult avoid.

Such alignment between the experimental and controlled environment of RCTs (and their selective research subjects) and the complexity of clinical practice (and their diverse and unique patients), is also reflected in – and facilitated by – trends further 'upstream' on this 'evidence chain', during the production of evidence. The reliance on evidence from observational studies has long been advocated as an alternative or addition to clinical trials (Black, 1996), but now the methodological designs of clinical trials are itself changing to better take into account the complex 'messiness' of practice and the differences between trial subjects and patients. Epstein (2007) has chronicled efforts of activists, policy makers and researchers to diversify evidence production by including more varied trial subjects so evidence is not only based on (and applicable to) the *l'homme moyen* (35 year old white male) but also produce evidence that is specifically based upon and applicable to other (sub)populations (women, children, elderly,

African-Americans). Epstein calls this “niche standardization”, and the “special populations” that become accepted as distinct ‘niches’ are social categories that have been categorized and measured in highly standardized ways (i.e. gender, ethnicity, age). Since their *biological* relevance is deemed plausible or acceptable by actors across political, medical, scientific and policy domains, these categories achieve the “categorical alignment” that allows them to travel and be agreed upon by heterogeneous actors for a range of (contradictory) purposes. Other social categories, that do not travel unquestionably and unaltered between biomedicine, state policy and social movements (i.e. sexual orientation, socio-economic status) remain too ‘political’ to become niches for trials. Other emerging research, such as ‘targeted’ cancer trials and pharmacogenomics, aim to identify genetic subcategories of patients that vary in their responses to treatments, promising a more ‘personalized’ medicine (Keating & Cambrosio, 2012: 376-381). Many other methodological innovations and modifications (cl)aim to better manage the ‘gap’ between the standardized ‘ideal conditions’ of RCTs and the diverse ‘real conditions’ in clinical practice now appear under the umbrella term ‘comparative effectiveness research’ (CER). CER has been characterized as “plac[ing] high value on external validity, or the ability to generalize results to real-world decision making” (Institute of Medicine, 2009: 38). This includes pragmatic trials that compare treatments within the ‘messiness’ of clinical practice, n-of-1-studies that compare treatments in individuals rather than populations, and the development of more ‘patient-centered’ outcome measures (Hoffman et al., unpublished). Social scientific inquiry that investigates the relations between CER and EBM and compares how they define the nature of evidence and trials are only just emerging (Luce et al., 2012; Hoffman et al., unpublished). One current CER initiative is using ‘implementation tools’ (such as HTA or guidelines) as the basis for recommending new research and trials. They do not recommend “designing gold-standard studies. It is understood that there are often tradeoffs required in designing studies that retain an adequate degree of methodologic rigor, while adopting features that make the results more generalizable to routine clinical practice.” (Tunis & Turkelson,

2012: 5). Thus, not only guidelines have lost their universal gold standard status, the 'universal evidence' of the traditional randomized clinical trial is also understood to be in need of modification to accommodate practices. This initiative illustrates that 'contextualization' is not limited to the application of evidence, but guidelines are used to design evidence production that is adapted to routine clinical practice. Local universality is not limited to the use of standards, nor to standard-setting, but is aimed for in trial design as well.

In all these efforts new hybrid categories, practices and objects are created that blur the boundaries between categories such as universal and local, the production and application of evidence, between science and practice. And even if the *ideals* of universal evidence, science or biology as well as locally diverse practice, culture and humans continue to persist as distinct categories, the hybrid practices and objects do change how these categories relate to one another. What you can *do* with 'universal evidence' and 'local practices' changes as the universal notions do not replace diversity, 'culture' or 'patient preferences', but the latter are acknowledged as legitimately (or at least inevitably) modifying 'universal' evidence and standards. So while few are (explicitly) aiming to break down the distinct categories altogether, in creating new cross-connections between different categories (e.g. non-evidentiary justification; local gold standards; objective values), the walls between categories may not be broken down entirely, neither are the same walls simply erected elsewhere (Busch, 2011: 5).

In conclusion, the meta-regulation of EBM guidelines adds a new chapter to the history of clinical practice guidelines. The norm of what 'good' guidelines are, has shifted from consensus statements produced by national experts, to universal gold standards determined by universal RCTs, to 'contextualized' recommendations regulated by universal standard-setting procedures. In shifting the determination of 'best practice' first away from individual clinicians, then from national experts, then from quantitative evidence, the notion of 'good' guidelines follows the transformations in the 'scientific' basis of medicine: "from a regime of trust in expertise and experts to a regime based on the mechanical generation of

data", to the most recent emergence of regulatory objectivity (Keating & Cambrosio, 2009: 325-6). In the latter regime 'best practice' is determined by procedures that regulate what counts as evidence, how to manage non-evidence and how (and which) experts are to act 'objectively'.

BIBLIOGRAPHY

AGREE Collaborative Group (2000) Guideline development in Europe. An international comparison. *International Journal of Technology Assessment in Health Care,* 16: 1039-49.

AGREE Collaboration (2003) Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Quality & Safety in Health Care*. 12: 18-23.

Akrich, M. (1992) The de-scription of technical objects. In W. E. Bijker and J. Law (Eds), *Shaping technology/building society. Studies in sociotechnical change* (pp. 205-24). Cambridge, MA: MIT Press.

Akrich, M., Leane, M., Roberts, C. and Nunes, J.A. (2012) Practising childbirth activism: a politics of evidence. *CSI Working Papers Series* N° 023. Downloaded on October 24, 2012 from: www.csi.ensmp.fr/Items/WorkingPapers/Download/DLWP.php?wp=WP_CSI_023.pdf

Akrich, M., Nunes, J., Paterson, F. & Rabeharisoa, V. (eds) (2008) *The Dynamics of Patient Organizations in Europe*. Paris: Presses des MINES.

Allen, D. (2009) From boundary concept to boundary object: The practice and politics of care pathway development. *Social Science & Medicine*, 69: 354-61.

Armstrong, D. (2002) Clinical autonomy, individual and collective: The problem of changing doctors' behaviour. *Social Science & Medicine,* 55: 1771-77.

Armstrong, D. (2007) Professionalism, Indeterminacy and the EBM Project. *BioSocieties*, 2(1): 73-84.

Barratt A. (2008) Evidence Based Medicine and Shared Decision Making: the challenge of getting both evidence and preferences into health care. *Patient Education & Counseling,* 73(3): 407-12

Barry, A. (2001) *Political Machines: Governing a Technological Society*. London: The Athlone Press.

Bastian H. (1996) Raising the standard: Practice Guidelines and Consumer Participation. *International Journal for Quality in Health Care*, 8(5): 485-90.

Bensing, J. (2000) Bridging the gap. The separate worlds of evidence-based medicine and patient-centered medicine. *Patient Education & Counseling,* 39: 17-25.

Berg, M. (1997) Problems and promises of the protocol. *Social Science & Medicine,* 44: 1081-88.

Berg, M. (1998) Order(s) and Disorder(s): of protocols and medical practices. In: M. Berg and A. Mol (eds.) *Differences in medicine. Unraveling Practices, Techniques, and Bodies* (Pp.226-246). Durham, NC: Duke University Press.

Berg, M., Horstman, K. Plass, S. and van Heusden, M. (2000) Guidelines, Professionals and the Production of Objectivity. Standardisation and the Professionalisation of Insurance Medicine. *Sociology of Health & Illness* 22: 765-91.

Berg, M., Ter Meulen, R., and van den Burg, M. (2001). Guidelines for appropriate care: the importance of empirical normative analysis. *Health Care Analysis*, 9(1): 77–99.

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

Black, D. (1998) The limitations of evidence. *Perspectives in Biology & Medicine* 42(1): 1.

Black N. (1996) Why we need observational studies to evaluate the effectiveness of health care. *BMJ,* 312: 1215–21.

Bloem, B.R., van Laar, T., Keus, S.H.J., de Beer, H., Poot, E., Buskens, E., Aarden, W., Munneke, M. on behalf of the Centrale Werkgroep Multidisciplinaire Richtlijn Parkinson 2006-2009 [Central Working Group Multidiscilpinary Guideline Parkinson 2006-2009]. (2010). *Multidisciplinaire richtlijn ziekte van Parkinson*. Alphen a/d Rijn: Van Zuiden Communications.

Blomgren M and Sahlin K (2007) Quests for transparency: Signs of a new institutional era in the health care field. In: T. Christensen & P. Lægreid (eds) *Transcending New Public Management*: *The Transformation of Public Sector Reforms* (155-177). Aldershot: Ashgate Publishing.

Borgerson, K. (2009) Valuing Evidence: Bias and the Evidence Hierarchy of Evidence-based Medicine. *Perspectives in Biology & Medicine* 52(2): 218-33.

Borgerson, K. (2005) Evidence-based Alternative Medicine? *Perspectives in Biology & Medicine*. 48(4): 502-15.

Bluhm, R. (2005) From hierarchy to network. A richer view of evidence for evidence

based medicine. *Perspectives in Biology & Medicine,* 48(4): 535-47.

Boivin, A. (2012) G-I-N PUBLIC toolkit introduction. In: *G-I-N PUBLIC Toolkit: Patient and Public Involvement in Guidelines* (Pp.10-14). Accessed, 05-12-2012 www.g-i-n.net/document-store/working-groups-documents/g-i-n-public/toolkit/toolkit-intro.pdf

Boivin, A., Currie, K., Fervers, B., Gracia, J., James, M., Marshall, C., Sakala, C., Sanger, S., Strid, J., Thomas, V., Van Der Weijden, T., Grol, R. and Burgers, J. (2010) Patient and public involvement in clinical guidelines: international experiences and future perspectives. *Quality & Safety in Health Care,* 19: e22.

Boivin, A., Green, J., Van Der Meulen, J., Légaré, F. and Nolte, E. (2009) Why consider patients' preferences? A discourse analysis of clinical practice guideline developers. *Medical Care,* 47: 908-15.

Boivin, A and Légaré, F. (2007) Public involvement in guideline production. [Letter]. *CMAJ* , 176(9): 1308-9.

Boivin, A., Légaré, F. and Gagnon, M. P. (2008) Competing norms: Canadian rural family physicians' perceptions of clinical practice guidelines and shared decision-making. *Journal of Health Services Research & Policy,* 13: 79-84.

Boivin, A., Légaré, F. and Lehoux, P. (2008) Decision technologies as normative instruments: exposing the values within. *Patient Education & Counseling,* 73: 426-30.

Boivin, A. and Marshall, C. (2008) What role for patients and the public in guidelines? Launch of a new G-I-N working group. *The EnGINe. Newsletter of the Guidelines International Network.* February, p.7. Accessed, 08-03-2013: http://www.g-i-n.net/document-store/engine/engine-february-2008

Boltanski, L., and Thévenot, L. (1999) The sociology of critical capacity. *European Journal of Social Theory*, 2: 359-78.

Böschen S, Kastenhofer K, Rust I, Soentgen J and Wehling P (2010) Scientific Nonknowledge and Its Political Dynamics: The Cases of Agri-Biotechnology and Mobile Phoning. *Science Technology & Human Values* 35: 783-811.

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

Bourret, P., and Rabeharisoa, V. (2008) Le jugement et la décision médicale en situation de forte incertitude: l'exemple de deux pratiques cliniques à l'épreuve de la génétique. *Sciences Sociales et Santé*, 26 : 33-66.

Bowker, G.C. and Star, S.L. (1996) How things (actor-net)work: Classification, magic and the ubiquity of standards. *Philosophia*, 25(3-4): 95-220

Bowker G.C and Star S.L. (2000) *Sorting things out. Classification and its consequences*. Cambridge, MA: MIT Press.

Briatte, F. (2012) The Politics of European Public Health Data. In: S. Greer and P. Kurzer (eds) *European Union Public Health Policy: Regional and global trends* (ch4). New York & London: Routledge.

Brody, H., Miller, F.G. and Bogdan-Lovis, E., (2005) Evidence-based medicine: watching out for its friends. *Perspectives in Biology & Medicine,* 48(4): 570–84.

Brouwers M., and Charette M. (2001) Evaluation of clinical practice guidelines in chiropractic care: A comparison of North American guideline reports. *Journal of the Canadian Chiropractic Association,* 45: 141-53.

Brouwers, M. C., Somerfield, M. R. and Browman, G. P. (2008) A for effort: learning from the application of the GRADE approach to cancer guideline development. *Journal of Clinical Oncology* 26: 1025-6.

Browman G, Levine M, Mohide E, Hayward R, Prtichard K, Gafni A, et al (1995) The practice guidelines development cycle: a conceptual tool for practice guidelines development and implementation. *Journal of Clinical Oncology*, 13: 502-512.

Brozek JL , Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW, Phillips B, Lelgemann M, Lethaby A, Bousquet J, Guyatt GH, Schünemann HJ and GRADE Working Group (2009) Grading quality of evidence and strength of recommendations in clinical practice guidelines. *Allergy,* 64(5): 669–677.

Buetow M and Kenealy T (2000) Evidence Based Medicine: The need for a new definition. *Journal of Evaluation in Clinical Practice* 6: 85-92.

Burgers, J., Cluzeau, F., Hanna, S., Hunt C., and Grol, R. (2003) Characteristics of high-quality guidelines. Evaluation of 86 Clinical Guidelines Developed in Ten European Countries and Canada. *International Journal of Technology Assessment in Health Care,* 19: 148-57.

Burgers, J. S., Bailey, J. V., Klazinga, N. S., Van Der Bij, A. K., Grol, R. and Feder, G. (2002) Inside guidelines: comparative analysis of recommendations and evidence in diabetes guidelines from 13 countries. *Diabetes Care,* 25: 1933-39.

Busch, L. (2011) *Standards. Recipes for Reality.* Cambridge, Mass.: MIT Press.

Callon, M. (1986) Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. In: John Law (ed) *Power, Action and Belief: A New Sociology of Knowledge* (pp.196-233) London: Routledge & Kegan Paul.

Cambrosio, A., Keating, P., Schlich, T. and Weisz, G. (2006) Regulatory objectivity and the generation and management of evidence in medicine. *Social Science & Medicine,* 63: 189-99.

Cambrosio, A., Keating, P., Schlich, T. and Weisz, G. (2009) Biomedical Conventions and Regulatory Objectivity: A Few Introductory Remarks. *Social Studies of Science,* 39(5): 651–64.

Caron-Flinterman, J., Broerse, J., and Bunders, J. (2005) The experiential knowledge of patients: a new resource for biomedical research? *Social Science & Medicine*, 60: 2575-84.

Cartwright, N. (2007) Are RCTs the Gold Standard? *BioSocieties* 2: 11-20.

Castel, P. (2002) *Normaliser les pratiques, organiser les médecins: la qualité comme stratégie de changement. Le cas des Centres de Lutte Contre le Cancer.* Doctoral Thesis. Institut d'Études politiques de Paris (Sociologie).

Castel, P. (2009) What's behind a guideline? Authority, Competition and Collaboration in the French Oncology Sector. *Social Studies of Science* 39(5): 743-64.

Castel P. and Merle, I. (2002) Quand les normes de pratiques deviennent une ressource pour les médecins. *Sociologie du Travail*, 44 : 337-55.

CBO (2007) *Evidence-Based Richtlijnontwikkeling. Handbook voor werkgroepleden*. [Evidence-Based Guideline Development. Manual for Working group members].] Accessed on sept 27th 2012 at: http://www.cbo.nl/Downloads/1250/EBRO_handleiding.pdf
(downloaded on: 26-05-2010).

CBO and NPCF (2009) *Blauwdruk patiëntenparticipatie in richtlijnontwikkeling. Leidraad voor richtlijnontwikkelaars en werkgroepleden*. [Blueprint patient

participation in guideline development. Guideline for guideline developers and working group members]. Utrecht: CBO Accessed on Sept 26th 2012 at: http://www.cbo.nl/thema/Richtlijnen/Pati-ntenparticipatie-voor-richtlijnontwikkelaar

Chong CAKY, Chen I, Naglie G, Krahn MD. (2009) How well do guidelines incorporate evidence on patient preferences? *Journal of General Internal Medicine,* 24(8): 977–82.

Church, J., Saunders, D., Wanke, M., Pong, R., Spooner, C. and Dorgan, M. (2002) Citizen participation in health decision-making: past experience and future prospects. *Journal of Public Health Policy,* 23: 12-32.

Cluzeau, F. (1998) *Strengthening effective practice: promoting the rigorous development of clinical guidelines in Europe through the creation of a common appraisal instrument.* EU BIOMED2 Programme application (BMH4-98-3669). Available: http://cordis.europa.eu/data/PROJ_BIOMED/ACTIONeqDndSESSIONeq22904200595ndDOCeq18ndTBLeqEN_PROJ.htm (Accessed 3 August 2009).

Cohen, J. (2004) Are clinical practice guidelines impartial? *International Journal of Technology Assessment in Health Care,* 20(4): 415-20.

Collins, H.M. ([1985] 1992) *Changing Order: Replication and Induction in Scientific Practice*. Chicago: University of Chicago Press.

Collins, H.M. and Evans, R.J. (2002) The Third Wave of Science Studies: Studies of Expertise and Experience. *Social Studies of Sciences,* 32(2): 235–296.

Council of Europe (2002) *Recommendation Rec(2001)13 on developing a methodology for drawing up guidelines on best medical practices*. Strasbourg: Council of Europe Publishing. Available: http://www.g-i-n.net/index.cfm?fuseaction=download&fusesub action=template&libraryID=50 (accessed 30 July 2009).

Crawford MJ, Rutter D, Manley C, Weaver T, Bhui K, Fulop N, et al. (2002) Systematic review of involving patients in the planning and development of health care. *BMJ,* 325: 1263-7

Cronje, R. and Fullan, A. (2003) Evidence-Based Medicine: Toward a New Definition of 'Rational' Medicine. *Health*, 7: 353-69.

Crozier, M. (1964) *The bureaucratic phenomenon*. Chicago: University of Chicago Press.

Daston, L. and Galison, P. (2007) *Objectivity*. Boston: Zone Books.

Delanty, G. and Rumford, C. (2005) *Rethinking Europe: Social Theory and the Implications of Europeanization.* (Ch8: pp.37-54). London: Routledge.

Detsky AS. (2006) Sources of bias for authors of clinical practice guidelines [editorial]. *CMAJ*, 175(9): 1033.

De Vries, R. and T. Lemmens (2006). The social and cultural shaping of medical evidence: Case studies from pharmaceutical research and obstetric science. *Social Science & Medicine*, 62(11): 2694–2706.

Díaz del Campo, P., Gracia, J., Blasco, J.A., Andradas, E. (2011) A strategy for patient involvement in clinical practice guidelines: methodological approaches. *BMJ Quality & Safety,* 20: 779-84.

Dobrow, M.J., Goel, V., Lemieux-Charles, L. and Black, N.A. (2006) The impact of context on evidence utilization: A framework for expert groups developing health policy recommendations. *Social Science & Medicine,* 63: 1811-24.

Dosquet, P., Goldberg, J., and Matillon Y. (1995) Comment élaborer des recommandations de pratique clinique? *Bulletin du Cancer*, *82*, 754-760.

EBRO (2007) *Handleiding voor werkgroepleden* [Manual for working group members]. Utrecht: Kwaliteitsinstituut voor de Gezondheidszorg CBO. Available at : http://www.cbo.nl/thema/Richtlijnen/EBRO-handleiding/PDF-versie (accessed 20 July 2012).

EBRO (not dated) *Formulier Vc voor het beoordelen van een Systematische Review van Observationeel Onderzoek* [Form Vc to assess systematic reviews of observational research]. Available at: http://www.cbo.nl/thema/Richtlijnen/EBRO-handleiding/G-Literatuurbeoordelings-formulieren/ (accessed: 12 April 2012).

Eccles, M., Clapp, Z., Grimshaw, J., Adams, P. C., Higgins, B., Purves, I., and Russell, I. (1996) North of England evidence based guidelines development project: methods of guideline development. *BMJ*, 312: 760-762.

Eddy, D. (1990) Practice policies: guidelines for methods. *JAMA*, 263: 839-41.

Eisenberg, J.M. (2002) Globalize the evidence, localize the decision: evidence-based medicine and international diversity. *Health Affairs,* 21: 166–68.

Epstein, S. (1995) The construction of lay expertise: AIDS activism and the forging of credibility in the reform of clinical trials. *Science Technology & Human Values*, 20(4): 408–37.

Epstein, S. (2007) *Inclusion: The Politics of Difference in Medical Research*, Chicago: University of Chicago Press.

Espeland, W. and Stevens, M. (2008) A Sociology of Quantification. *European Journal of Sociology (Archives Européennes de Sociologie),* 49: 401-36.

Evidence Based Medicine Working Group (1992) Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*, 268: 2420-5.

Fédération Nationale des Centres de Lutte Contre le Cancer (1994), *Rapport d'activité*. Paris: FNCLCC.

Feldman, M. S., and Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, 48: 94-118.

Fervers, B., Burgers, J. S., Haugh, M. C., Latreille, J., Mlika-Cabanne, N., Paquet, L., Coulombe, M., Poirier, M., and Burnand, B. (2006). Adaptation of clinical guidelines: literature review and proposition for a framework and procedure. *International Journal for Quality in Health Care*, 18: 167-76.
Fervers, B., Hardy, J., and Philip, T. (Eds.) (2001). Standards, Options and Recommendations. Clinical Practice Guidelines for cancer care from the French National Federation of Cancer Centres (FNCLCC). *British Journal of Cancer*, 84(Suppl 2): 1-92.

Fervers, B., Remy-Stockinger, M., Mazeau-Woynar, V., Otter, R., Liberati, A., Littlejohns, P., Qureshi, S., Vlayen, J., Characiejus, D., Corbacho, B., Garner, S., Hamza-Mohamed, F., Hermosilla, T., Kersten, S., Kulig, M., Leshem, B., Levine, N., Ballini, L., Middelton, C., Mlika-Cabane, N., Paquet, L., Podmaniczki, E., Ramaekers, D., Robinson, E., Sanchez, E. and Philip, T. (2008) CoCanCPG. Coordination of cancer clinical practice in Europe. *Tumori,* 94: 154-59.

Field, M.J., and Lohr, K.N. (Eds). (1990) *Clinical Practice Guidelines: Directions for a New Program, Institute of Medicine*, Washington, DC: National Academy Press.

Florin, D. and J. Dixon (2004) Public involvement in health care. *BMJ*, 328: 159–161.

Forbat, L., Hubbard, G. and Kearney, N. (2009) Patient and public involvement: models and muddles. *Journal of Clinical Nursing,* 18: 2547-54.

Foucault, M. (1994) *Dits et écrits 1954-1988. Vol. III 1976-1979*. Paris: Gallimard.

Fox, R. (2003) Medical Uncertainty Revisited. In G. L. Albrecht, R. Fitzpatrick, and S. C. Scrimshaw (Eds), *Handbook of social studies in health and medicine* (pp. 409-25). Thousand Oaks, CA: Sage.

Fowler, P.B.S. (1997) Evidence Based Everything *Journal of Evaluation of Clinical Practice* 3(3): 239-43.

Fried, M., Farthing, M., Krabshuis, J., Quigley, E. on behalf of the WGO-OMGE Global Guidelines Task Force (2006) Global guidelines: is gastroenterology leading the way? *Lancet*, 368: 2041-42

Freidson, E. (1970). *Profession of medicine*. New York: Dodd, Mead.

Freidson, E. (1984) *The changing nature of professional control. Annual Review of Sociology*, 10: 1-20.

Garsten, C. and Lindh De Montoya, M. (2008) Introduction: examining the politics of transparency. In: C. Garsten and M. Lindh de Montoya (eds) *Transparency in a new global order : unveiling organizational visions* (pp. 1-21). Cheltenham: Edward Elgar.

Goldenberg, M.J. (2006) On evidence and evidence-based medicine: lessons from the philosophy of science. *Social Science & Medicine,* 62(11): 2621-32.

Goldenberg, M.J. (2009) Iconoclast or Creed? Objectivism, Pragmatism, and the Hierarchy of Evidence. *Perspectives in Biology & Medicine*, 52: 168-87.

Goldenberg, M.J., Borgerson, K. and Bluhm R. (2009) The Nature of Evidence in Evidence Based Medicine. Guest editors' introduction. *Perspectives in Biology & Medicine*, 52: 164-7.

Goodman, K.W. (2003) *Ethics and evidence-based medicine: Fallibility and responsibility in clinical science*. Cambridge: Cambridge University Press.

GRADE working group. (2004) Systems for grading the quality of evidence and the strength of recommendations I: Critical appraisal of existing approaches. *BMC Health Services Research*, 4: 38.

GRADE Working Group, Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, et al. (2004) Grading quality of evidence and strength of recommendations. *BMJ* 328: 1490-1494.

Graham I.D., Beardall S., Carter A.O., Glennie J., Hébert P.C., Tetroe J.M., McAlister F.A., Visentin S. and Anderson G.M. (2001) What is the quality of drug therapy clinical practice guidelines in Canada?. *CMAJ,* 165: 157-63.

Greene, J. (2007) *Prescribing by Numbers. Drugs and the definition of disease*. Baltimore: John Hopkins University Press.

Greer, S.L. (2006) *Responding to Europe. Government, NHS and stakeholder responses to the EU health policy challenge,* London: The Nuffield Trust.

Greer, A.L., Goodwin, J.S., Freeman, J.L., and Wu, Z.H. (2002). Bringing the patient back in. Guidelines, practice variations, and the social context of medical practice. *International Journal of Technology Assessment in Health Care,* 18(4): 747-61.

Grol, R., Cluzeau, F. A. and Burgers, J. S. (2003) Clinical practice guidelines: towards better quality guidelines and increased international collaboration. *British Journal of Cancer,* 89 Suppl 1: S4-8.

Grol R, Eccles M, Maisonneuve H, Woolf S. (1998) Developing clinical practice guidelines: the European experience*. Disease Management and Health Outcomes,* 4: 255-66.

Grossman, J. and Mackenzie, F.J. (2005) The randomized controlled trial: gold standard, or merely standard? *Perspectives in Biology & Medicine,* 48: 516-34.

Guidelines International Network (2002) *Position Paper*, Available: http://www.g-i-n.net/about-g-i-n/history-of-g-i-n (accessed 22 March 2010).

G-I-N PUBLIC Steering Committee (2012) Forword. In: *G-I-N PUBLIC Toolkit: Patient and Public Involvement in Guidelines* (Pp.5-8). Available: www.g-i-n.net/document-store/working-groups-documents/g-i-n-public/toolkit/toolkit-combined.pdf (accessed, 05 March 2013)

Guyatt GH et al. (2000) Practitioners of evidence based care. Not all clinicians need to appraise evidence from scratch but all need some skills. *BMJ* , 320: 954–955.

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P. and Schunemann, H. J. (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ,* 336: 924-6.

Hannes K, Van Royen P, Aertgeerts B, Buntinx F, Ramaekers D, and Chevalier P. (2005) La validation systémique de guides de pratique clinique: l'instrument

AGREE, elaboration et diffucion d'une liste internationale de critères. *Revue Médicale de Liège*, 60: 949-56.

Hayward RSA, Wilson MC, Tunis SR, Bass EB, Guyatt G, for the Evidence-Based Medicine Working Group. (1995) Users' guides to the medical literature VIII. How to use clinical practice guidelines. Are the recommendations valid? *JAMA*, 274: 570-4.

Healy, D. (2009) Trussed in Evidence? Ambiguities at the Interface between Clinical Evidence and Clinical Practice. *Transcultural Psychiatry*; 46(1): 16-37

Hedgecoe, A. (2002). Reinventing diabetes: Classification, division, and the geneticization of disease. *New Genetics & Society,* 21(1): 7–27.

Hedgecoe, A. (2003) Expansion and uncertainty: Cystic fibrosis, classification and genetics. *Sociology of Health & Illness*, 25: 50-70.

Hervey, T. K. (2008) The European Union's governance of health care and the welfare modernization agenda. *Regulation & Governance,* 2: 103-120.

Hirsch, J., and Guaytt, G. (2009) 'Clinical experts or methodologists to write clinical guidelines?' *The Lancet,* 374: 273-5.

Hoeyer, K. (2010) An anthropological analysis of European Union (EU) health governance as biopolitics: the case of the EU tissues and cells directive. *Social Science & Medicine,* 70: 1867-73.

Hoffman, A.S., Cambrosio, A. and Battista, R. (unpublished) The Use of Health Technology Assessment in Comparative Effectiveness Research.

Hogle, L. (1995) Standardization across non-standard domains: The case of organ procurement. *Science Technology & Human Values,* 20: 482-501.

Hogle, L. (2009) Pragmatic Objectivity and the Standardization of Engineered Tissues. 2009. *Social Studies of Science*. 39(5): 1-26.

Hughes, D. Mullen, C. Vincent-Jones, P. (2009) Choice vs. voice? PPI policies and the re-positioning of the state in England and Wales. *Health expectations*. 12: 237–250.

Hurdowar, A., Graham, I.D., Bayley, M., Harrison, M., Wood-Dauphinee, S., and Bhogal S. (2007) Quality of stroke rehabilitation clinical practice guidelines. *Journal of Evaluation in Clinical Practice,* 13: 657-64.

Hutchings, A. and Raine, R. (2006). A systematic review of factors affecting the judgments produced by formal consensus development methods in health care. *Journal of Health Services Research & Policy*, 11: 172-179.

Institute of Medicine (2009) *Initial National Priorities for Comparative Effectiveness Research*. Washington, DC: The National Academies Press.

Institute of Medicine. (2011) *Clinical Practice Guidelines We Can Trust*. Washington, DC: National Academies Press.

Jacoby, I., (1993). Consensus Development at NIH: What Went Wrong? *Risk. Issues in Health & Safety* 4:133-42.

Jasanoff, S. (2003) Breaking the Waves in Science Studies: Comment on H.M. Collins and Robert Evans, The Third Wave of Science Studies. *Social Studies of Science,* 33(3): 389-400.

Kahan, J.P., Park, R.E., Leape, L.L., Bernstein, S.J., Hilborne, L.H., Parker, L., Kamberg, C.J., Ballard, D.J. and Brook, R.H. (1996) Variations by specialty in physician ratings of the appropriateness and necessity of indications for procedures. *Medical Care*. 34(6): 512-23.

Keating, P. and Cambrosio, A. (2009) Who's minding the data? Data Monitoring Committees in clinical cancer trials. *Sociology of Health & Illness,* 31: 325-42.

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

Kelson, N. (2001) Patient involvement in clinical guideline development - where are we now? *Journal of Clinical Governance* 9: 169-74.

Knaapen, L. (in press) Evidence Searched Guidelines: How to manage an absence of evidence in Evidence Based Medicine. *Social Studies of Science*

Knaapen, L. (2012) European regulation and harmonization of Clinical Practice Guidelines". In: Scott Greer and Paulette Kurzer (eds) *European Union Public Health Policy: Regional and global trends* (pp.64-80). New York & London: Routledge.

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

Knorr-Cetina K (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.

Krahn, M., and Naglie, G. (2008) The next step in guideline development. Incorporating patient preferences. *JAMA*, 300: 436-8.

Krahn, M. (2010) 'New' Evidence for Clinical Practice Guidelines: Should we Search for 'Preference Evidence'? *The Patient: Patient-Centered Outcomes Research* 3(2): 71-77.

Lambert, H. (2006) Accounting for EBM: notions of evidence in medicine. *Social Science & Medicine,* 62: 2633-45.

Lampland, M. and Star, S. L. (2009) *Standards and their stories : how quantifying, classifying, and formalizing practices shape everyday life,* Ithaca: Cornell University Press.

Latimer J, Heatherstone K, Atkinson P, Clarke A, Pilz DT and Shaw A. (2006) Rebirthing the Clinic: The Interaction of Clinical Judgment and Genetic Technology in the Production of Medical Science. *Science Technology & Human Values,* 31: 599-630.

Latour, B. (1983) Give Me a Laboratory and I will Raise the World. In K. Knorr-Cetina and M. Mulkay (eds.) *Science Observed: Perspectives on the Social Study of Science* (pp.141-70) London and Beverly Hills: Sage.

Latour, B. (1987) *Science in Action*. Cambridge: Harvard University Press.

Latour, B. (1990). Drawing things together. In M. Lynch and S. Woolgar (Eds.), *Representation in scientific practice* (pp. 19-68). Cambridge: MIT Press.

Latour, B. (2007) *Reassembling the Social. An introduction to Actor-Network-Theory*. New York: Oxford University Press.

Leape, L.L., Park, R.E., Kahan, J.P., and Brook, R.H. (1992) Group judgments of appropriateness: the effect of panel composition. *Quality Assurance in Health Care*. 4(2): 151-9.

Learmonth, M., Martin, G.P. and Warwick, P. (2009). Ordinary and effective: The catch-22 in managing the public voice in health care? *Health Expectations*, 12: 106-115.

Légaré, F, Boivin, A., van der Weijden, T., Pakenham, C., Burgers, J., Légaré, J., St-Jacques, S. and Gagnon, S. (2011) Patient and Public Involvement in Clinical

Practice Guidelines: Knowledge Synthesis of Existing Programs. *Medical Decision Making* 31(6): 45-74.

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

Lehoux, P., Daudelin, G., Demers-Payette, O., and Boivin, A. (2009). Fostering deliberations about health innovations : What do we want to know from the publics? *Social Science and Medicine*. 68(11): 2002-09.

Levay, C. and Waks, C. (2009) Professions and the Pursuit of Transparency in Healthcare: Two Cases of Soft Autonomy. *Organization Studies* 30(5): 509-527.

Light, D. (2000) The Medical Profession and Organizational Change: From Professional Dominance to Countervailing Power. In: C. Bird, P. Conrad, A. Fremont (Eds) *Handbook of Medical Sociology* (Pp.201-17) Upper Saddle River, NJ: Prentice Hall.

Lohr, K. and Field, M. eds. (1992) *A provisional instrument for assessing clinical practice guidelines. In Guidelines for Clinical Practice: From Development to Use.* Washington DC: National Academy Press.

Luce, B.R., Drummond, M., Jönsson, B., Neumann, P.J, Schwartz, J .S., Siebert, U. and Sullivan, S.D. (2012) EBM, HTA, and CER: Clearing the Confusion. *The Milbank Quarterly*, 88(2): 256–276

MacDermid, J.C., Brooks, D., Solway, S., Switzer-McIntyre, S., Brosseau, L. and Graham, I.D. (2005) Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines, *BMC Health Services Research*, 5: 18.

Manchikanti, L., Singh, V., Derby, R., Helm, S., Trescot, A.M., Staats, P.S., Prager, J.P. and Hirsch, J.A. (2008) Review of Occupational Medicine Practice Guidelines for Interventional Pain Management and Potential Implications. *Pain Physician,* 11: 271-89.

Martin, G. P. (2008). “Ordinary people only”: knowledge, representativeness, and the publics of public participation in healthcare. *Sociology of Health & Illness*, 30(1): 35-54.

May, C. (2006) Mobilising modern facts: Health technology assessment and the politics of evidence. *Sociology of Health & Illness*, 28: 513-32.

McGivern, G. and Fischer, M.D. (2012) Reactivity and reactions to regulatory transparency in medicine, psychotherapy and counselling. *Social Science & Medicine* 74(3): 289-296.

McGoey, L. (2009) Pharmaceutical Controversies and the Performative Value of Uncertainty. *Science as Culture,* 18: 151-164.

McGuire, W.L. (2005) Beyond EBM: new directions for evidence-based public health. *Perspectives in Biology & Medicine,* 48(4): 557-69.

Michie, S., and Lester, K. (2005) Words matter: increasing the implementation of clinical guidelines. *Quality & Safety in Health Care*, 14: 367–370.

Mol, A. (2006) *De logica van het zorgen; Actieve patiënten en de grenzen van het kiezen*. Amsterdam: Van Gennep

Monaghan, M. (2008) The evidence base in UK drug policy: the new rules of engagement. *Policy and Politics* 36: 145-150.

Moreira, T. (2005) Diversity in Clinical Guidelines: The Role of Repertoires of Evaluation. *Social Science & Medicine* 60(9): 1975–85.

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

Moreira, T. (2012) Health Care Standards and the Politics of Singularities : Shifting In and Out of Context. *Science Technology & Human Values,* 37(4): 306-31.

Moreira, T., May, C. and Bond, T. (2009) Regulatory Objectivity in Action. Mild Cognitive Impairment and the Collective Production of Uncertainty. *Social Studies of Science* 39(5): 665-690.

Moreira, T., May, C., Mason, J. and Eccles, M. (2006) A new method of analysis enabled a better understanding of clinical practice guideline development processes. *Journal of Clinical Epidemiology*, 59(11): 1199-206.

Murphy JF. (2008) Paternalism or partnership: clinical practice guidelines and patient preferences. *Irish Medical Journal*, 101: 232.

Mykhalovskiy, E. (2003). Evidence-based medicine: Ambivalent reading and the clinical recontextualization of science. *Health*, 7(3): 331–352.

Mykhalovskiy, E., and Weir, L. (2004) The problem of evidence-based medicine: Directions for social science. *Social Science & Medicine*, 59: 1059-1069.

Nes, S. and Moen, A. (2010) Constructing standards - A study of nurses negotiating with multiple modes of knowledge. *Journal of Workplace Learning* 22(6): 376- 393.

Moen, A. and Nes, S. (2012) Consolidating work descriptions: Creating shared knowledge objects. In: A. Moen, A. Mørch, and S. Paavola (Eds.) *Collaborative Knowledge Creation. Practices, tools, concepts*. (pp. 249-260) Rotterdam: Sense Publishers.

Nilsen, E. S., Myrhaug, H. T., Johansen, M., Oliver, S., and Oxman, A. D. (2006). Methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material (review), *Cochrane Database of Systematic Reviews*, issue 3. Art. CD004563.

Nuckols, T. K., Lim, Y. W., Wynn, B. O., Mattke, S., Maclean, C. H., Harber, P., Brook, R. H., Wallace, P., Garland, R. H. and Asch, S. (2008) Rigorous development does not ensure that guidelines are acceptable to a panel of knowledgeable providers. *Journal of General Internal Medicine,* 23: 37-44.

O'Connell, J. (1993) Metrology: The Creation of Universality by the Circulation of Particulars. *Social Studies of Science*, 23: 129-73.

Ollenschläger G, Marshall C, Qureshi S, *et al.* (2004) Improving the quality of health care: using international collaboration to inform guidelines programmes by founding the Guidelines International Network (G-I-N). *Quality & Safety in Health Care*, 13: 455–60.

Pagliari, C. and Grimshaw, J. (2002) Impact of group structure and process on multidisciplinary evidence-based guideline development: An observational study. *Journal of Evaluation in Clinical Practice*, 8: 145-153.

Pentheroudakis, G. Stahel, R., Hansen, H. and Pavlidis, N. (2008) Heterogeneity in cancer guidelines: should we eradicate or tolerate? *Annals of Oncology* 19: 2067–2078

Pinell, P. (2002) *The fight against cancer: France 1890-1940*. London: Routledge.

Pope, C. (2003) Resisting evidence: the study of evidence-based medicine as a contemporary social movement. *Health*, 7: 267-282

Porter, T.M. (1995). *Trust in numbers: the pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Power, M. (1999) *The Audit Society: Rituals of Verification*, Oxford: Oxford University Press.

Power, M. (2000) The Audit Society – Second Thoughts. *International Journal of Auditing* 4: 111-119.

Prior, Lindsay. (2003) Belief, knowledge and expertise: the emergence of the lay expert in medical sociology. *Sociology of Health and Illness,* 25: 43-57.

Program in Evidence-Based Care (2007) *Proceedings of the 1st Scientific Symposium "The Nature of Evidence".* 28 February 2007 Toronto. Available at: https://www.cancercare.on.ca/common/pages/UserFile.aspx?fileId=39496 (accessed 12 May 2012)

Qaseem, A., Forland, F., Macbeth, F., Ollenschläger, G., Phillips, S., van der Wees, P. and Board of Trustees of the Guidelines International Network (2012) Guidelines International Network: toward international standards for clinical practice guidelines. *Annals of Internal Medicine* 156(7): 525-531.

Raats, I. (2009) Rol van patienten bij richtlijnontwikkeling [role of patients in guideline development]. Presentation at 'Werkconferentie richtlijnen en geneesmiddelen', Coordinatiegroep geneesmiddelenbeleid (NFK, CG-raad, NPCF), 18 juni 2009.

Raats CJ, van Veenendaal H, Versluijs MM, Burgers JS. (2008) A generic tool for development of decision aids based on clinical practice guidelines. *Patient Education & Counseling,* 73(3): 413-417.

Rabeharisoa, V. (2006) From representation to mediation: The shaping of collective mobilization on muscular dystrophy in France. *Social Science & Medicine*, 62(3): 564-576.

Rabeharisoa, V. and Bourret, P. (2009) Staging and Weighting Evidence in Biomedicine: Comparing Clinical Practices in Cancer Genetics and Psychiatric Genetics. *Social Studies of Science* 39(5): 691-715.

Rabeharisoa, V. and Callon, M. (2004) Patients and Scientists in French Muscular Dystrophy Research, In Jasanoff, S., (ed.), *States of Knowledge. The Co-Production of Science and Social Order* (pp.142-160). London and New York, Routledge.

Raine, R., Sanderson, C., Hutchings, A., Carter, S., Larkin, K. and Black. N. (2004) An experimental study of determinants of group judgments in clinical guideline development. *Lancet*. 6; 364(9432): 429-37.

Reynaud, J.D. and Richebé, N. (2007) Règles, conventions et valeurs. Plaidoyer pour la normativité ordinaire. *Revue française de sociologie*, 48(1): 3-36.

Rogers, W. A. (2002) Evidence-based medicine in practice: limiting or facilitating patient choice? *Health Expectations* 5: 95-103.

Rumford, C. (2002) *The European Union: a political sociology* (pp.46-81). Oxford: Blackwell Publishing.

Sackett, D.L. et al. (2000). *Evidence-based medicine : how to practice and teach EBM*. Edinburgh ; New York: Churchill Livingstone.

Sackett, D. L. and Rosenberg, W. M. (1995) The need for evidence-based medicine. *Journal of the Royal Society of Medicine,* 88: 620-4.

Schünemann, H.J., Fretheim, A. and Oxman, A.D. (2006) Improving the use of research evidence in guideline development: 10. Integrating values and consumer involvement. *Health Research Policy & Systems,* 5: 4-22.

Scott, J. and Trubek, D.M. (2002) Mind the gap: law and new approaches to governance in the European Union. *European Law Journal,* 8: 1-18.
Smith B. et al. (1996) Evidence Based Medicine [Letters to the Editor] *BMJ*, 313(7050): 169-171.

Shaneyfelt, T. (2012) In guidelines we cannot trust. *Archives of Internal Medicine*, 172(21): 1633-34.

Star, S. L. and J. R. Griesemer. (1989) Institutional Ecology, Translations, and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907 - 1939. *Social Studies of Science* 19: 387-420.

Tambuyzer E, Pieters G, Van Audenhove C. (2011) Patient involvement in mental health care: one size does not fit all. *Health Expectations*. Epub ahead of print. doi: 10.1111/j.1369-7625.2011.00743.x

Timmermans, S. (2005) From autonomy to accountability: the role of clinical practice guidelines in professional power. *Perspectives in Biology & Medicine* 48(4): 490-502.

Timmermans, S. (2011) The Joy of Science: Finding Success in a "Failed" Randomized Clinical Trial. *Science, Technology & Human Values* 36: 549-572.

Timmermans, S. and Berg, M. (1997) Standardization in action: Achieving local universality through medical protocols. *Social studies of science* 27: 273-305.

Timmermans, S. and Berg, M. (2003) *The Gold Standard: A challenge of Evidence-Based Medicine and Standardization in Health Care*. Philadelphia: Temple University Press.

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

Timmermans, S. and Kolker, E. (2004) Evidence-Based Medicine and the Reconfiguration of Medical Knowledge. *Journal of Health and Social Behavior*, 45: 177-193.

Timmermans, S. and Mauck, A. (2005) The promises and pitfalls of Evidence-Based Medicine. *Health Affairs,* 24: 18-28.

Topalian, J. and Pagliari, C. (2007) *Exploring international variation in consumer involvement in guideline development*. Presentation at Annual G-I-N Conference 2007, Toronto, Canada.

Tonelli, M.R. (2006) Integrating evidence into clinical practice: An alternative to evidence-based approaches. *Journal of Evaluation in Clinical Practice* 12(3): 248–256.

Tritter, J.Q. (2009) Revolution or evolution: the challenges of conceptualizing patient and public involvement in a consumerist world. *Health Expectations* 12(3): 275-87.

Trubek, L.G., Rees, J.V., Hoflund, A.B., Farquhar, M., and Heimer, C.A. (2008) Health care and new governance: The quest for effective regulation – Introduction, *Regulation & Governance*, 2: 1-8.

Tunis, S.R. and Turkelson, C. (2012) Using Health Technology Assessment to Identify Gaps in Evidence and Inform Study Design for Comparative Effectiveness Research *Journal of Clinical Oncology* Epub ahead of print. Doi: 10.1200/JCO. 2012.42.6338

Umscheid C.A. (2009) Should guidelines incorporate evidence on patient preferences? *Journal of General Internal Medicine* 24: 988-90.

Van de Bovenkamp, H.M. and Trappenburg M.J. (2009) Reconsidering Patient Participation in Guideline Development. *Health Care Analysis*; 17: 198-216.

Van de Bovenkamp, H.M., Grit, K.J. and Bal, R.A., (2008). Inventarisatie patiëntenparticipatie in onderzoek, kwaliteit en beleid. Rotterdam: Instituut Beleid en Management Gezondheidszorg. Available from: http://oldwww.bmg.eur.nl/personal/r.bal/vd%20Bovenkamp%20et%20al%20-%202008%20-%20inventarisatie%20patientenparticipatie.pdf (accessed September 26, 2012)

Van der Wees, P.J., Hendriks, E.J., Custers, J.W., Burgers, J.S., Dekker, J. and Bie R.A. de (2007) Comparison of international guideline programs to evaluate and update the Dutch program for clinical guideline development in physical therapy. *BMC Health Servi*ces Research, 23: 191.

Van der Weijden T, Légaré F, Boivin A, Burgers JS, van Veenendaal H, Stiggelbout AM, Faber M and Elwyn G. (2010) How to integrate individual patient values and preferences in clinical practice guidelines? A research protocol. *Implementation Science,* 5: 10.

Van Veenendaal, H., Franx, G.C., Grol, M.H., van Vuuren, J., Versluijs, M. and Dekhuyzen, P.N.R. (2004). Patiëntenparticipatie bij richtlijnontwikkeling. In: Van Everdingen, J.J.E., Burgers, J.S., Assendelft, W.J.J. et al. (eds) *Evidence-based richtlijnontwikkeling* (pp.48-62). Houten: Bohn Stafleu Van Loghum.

Van Wersch A and Eccles M. (2001) Involvement of consumers in the development of evidence based clinical guidelines: practical experiences from the North of England evidence based guideline development programme. *Qual Health Care* 10(1): 10-16.

Verkerk K, Van Veenendaal H. Severens JL, Hendriks EJ and Burgers JS (2006) Considered judgement in evidence-based guideline development. *International Journal for Quality in Health Care,* 18: 365-369.

Vlayen, J., Aertgeerts, B., Hannes, K., Sermeus, W. and Ramaekers, D. (2005) A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *International Journal for Quality in Health Care,* 17: 235-42.

Wait, S., and Nolte, E. (2006). Public involvement policies in health: exploring their conceptual basis. *Health Economics, Policy and Law*, 10(2): 149-162

Watine, J., Friedberg, B., Nagy, E., Onody, R., Oosterhuis, W., Bunting, P. S., Charet, J. C. and Horvath, A. R. (2006) Conflict between guideline methodologic quality and recommendation validity: a potential problem for practitioners. *Clinical Chemistry,* 52: 65-72.

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

Wennberg, J., and A. Gittelsohn (1973) Small Area Variations in Health Care Delivery. *Science* 182(117): 1102–8.

Will, C.M. (2009). Identifying Effectiveness in "The Old Old": Principles and Values in the Age of Clinical Trials. *Science Technology & Human Values,* 34: 607-628.

Will, C.M. (2010) Addressing the commercial context: the collective critique of clinical trials. In: C. Will and T. Moreira (eds) *Medical proofs, social experiments: Clinical trials in shifting contexts* (pp.67-83). Farnham, UK: Ashgate Publishers.

Wynne, B. (1992) Misunderstood Misunderstandings: Social Identities and Public Uptake of Science. *Public Understanding of Science* 1: 281-304.

Wynne, B. 2003) Seasick on the Third Wave? Subverting the Hegemony of Propositionalism: Response to Collins and Evans (2002). *Social Studies of Science*, 33(3): 401–417.

Zitzelsberger, L., Grunfeld, E. and Graham, I. (2004) Family physicians' perspectives on practice guidelines related to cancer control. *BMC Family Practice*, 5: 15-25.

Zuiderent-Jerak, T. (2007) Preventing implementation: exploring interventions with standardization in healthcare. *Science as Culture*, 16: 311-29.

Zuiderent-Jerak, T. (2009) Competition in the Wild: Reconfiguring Healthcare Markets. *Social Studies of Science* 39(5): 765-792.

Zuiderent-Jerak, T. and Bruun Jensen, C. (2007) Editorial Introduction: Unpacking 'Intervention' in Science and Technology Studies. *Science as Culture* 16(3): 227-235.

Zuiderent-Jerak, T., Forland. F., and Macbeth F. (2012) Guidelines should reflect all knowledge, not just clinical trials. *BMJ*, 345: e6702.