INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

SELECTING SUBJECTS FOR PARTICIPATION IN CLINICAL RESEARCH: AN EMPIRICAL INQUIRY AND ETHICAL ANALYSIS

Charles Weijer
B.Med.Sc., M.D., B.A.(*Hons. Phil.*), M.Sc.(*Bioethics*)
Division of Experimental Medicine
Faculty of Medicine
McGill University, Montreal

May 1997

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Doctor of Philosophy in Experimental Medicine.

Abstract (English)

Procedures for the selection of subjects for participation in randomized clinical trials -- usually formalized as eligibility criteria in the study protocol -- have both scientific and ethical implications. In this thesis, I undertake an examination of eligibility criteria at three stages in the genesis and dissemination of medical knowledge: clinical trial protocol, interpretation by investigators, and reporting of study results.

In the first chapter, ethical issues in subject selection are reviewed and the main study questions are presented. In the second chapter, the results of an examination of eligibility criteria in two sets of clinical trials, one sponsored by the NSABP, the other sponsored by POG, covering a twenty-year time span are presented. The POG trials had far fewer eligibility criteria than the NSABP studies, suggesting that large numbers of criteria may not be necessary for high quality research. In the third chapter, the impact of subjective eligibility criteria on enrollment and investigator uncertainty is explored. Subjective criteria were associated with more variable enrollment decisions and greater uncertainty. Such criteria represent a threat to the validity, conduct and interpretation of trials and, therefore, should only be included when carefully justified. The fourth chapter examines the *accuracy* of the reporting of eligibility criteria in sets of corresponding study protocol, methods paper, journal article, and Clinical Alert.

Important information is lost at each step in the dissemination of study results. Unnecessary criteria ought to be dropped at a trial's inception; all other criteria must be reported faithfully. The fifth chapter attempts to provide a comprehensive philosophical account of just selection procedures for clinical research using the political philosophy of Michael Walzer. The sixth, and last, chapter, discusses explanatory and pragmatic approaches to clinical trial design, overlapping scientific and ethical concerns related to eligibility criteria, and questions for further study.

Abstract (French)

Les procédures de sélection des sujets invités à participer à des essais cliniques randomisés (qui s'inscrivent généralement dans le cadre des critères d'admissibilité des protocoles de recherche) revêtent des conséquences à la fois scientifiques et éthiques. Dans cette thèse, j'examine les critères d'admissibilité à trois étapes de la genèse et de la divulgation des connaissances médicales : protocole d'essai clinique, interprétation par les chercheurs et établissement de rapports sur les résultats de la recherche.

Le premier chapitre examine les questions éthiques liées à la sélection des sujets et présente les principales questions à l'étude. Le deuxième chapitre est consacré aux résultats d'un examen des critères d'admissibilité dans deux ensembles d'essais cliniques, les uns parrainés par le *NSABP* et les autres par le *POG*, sur une période de vingt ans. Les essais *POG* ont beaucoup moins de critères d'admissibilité que les essais *NSABP* ce qui donne à penser qu'un nombre important de critères n'est pas nécessairement un gage de la qualité de la recherche. Le troisième chapitre s'intéresse à l'impact des critères d'admissibilité subjectifs sur la participation des sujets et l'incertitude des chercheurs. Les critères subjectifs sont associés à des décisions en matière de participation plus variables et à une plus grande incertitude. Ces critères menacent la validité, le déroulement et l'interprétation des essais et partant, ne devraient être inclus qu'au prix d'une solide

justification. Le quatrième chapitre examine avec quelle *exactitude* les critères d'admissibilité sont signalés dans les ensembles correspondants de protocoles, dans les articles sur les méthodes, dans les articles qui paraissent dans des revues savantes et dans *Clinical Alert*. D'importantes données se perdent à chaque étape de la divulgation des résultats de l'étude. Les critères inutiles devraient être exclus lors de la conception de l'essai; tous les autres critères doivent faire l'objet d'une signalisation fidèle. Le cinquième chapitre tente de fournir un compte rendu philosophique complet des procédures de sélection visant la recherche clinique sur la base de la philosophie politique de Michael Walzer. Le sixième et dernier chapitre présente les méthodes explicatives et pragmatiques de conception des essais cliniques, en recoupant les préoccupations scientifiques et éthiques liées aux critères d'admissibilité et soulève des questions auxquelles d'autres études pourront éventuellement tenter de répondre.

Preface

In accordance with the Faculty of Graduate Studies and Research document "Guidelines for Thesis Preparation" the candidate has taken the option, according to section 3, of writing the experimental part of the thesis (chapters two through five) in the form of original papers submitted for publication to learned journals. This provision reads as follows:

> Candidates have the option of including, as a part of the thesis, the text of one or more papers submitted or to be submitted for publication, or the clearly-duplicated text of one or more published papers. These texts must be bound as an integral part of the thesis.
>
> If this option is chosen, **connecting texts that provide logical bridges between the different papers are mandatory**. The thesis must be written in such a way that it is more than a mere collection of manuscripts; in other words, results of a series of papers must be integrated.
>
> The thesis must still conform to all other requirement of the "Guidelines for Thesis Preparation". **The thesis must include**: A Table of Contents, an abstract in English and French, an introduction

which clearly states the rationale and objectives of the study, a review of the literature, a final conclusion and summary, and a thorough bibliography or reference list.

Additional material must be provided where appropriate (e.g. in appendices) and in sufficient detail to allow clear and precise judgement to be made of the importance and originality of the research reported in the thesis.

In the case of manuscripts co-authored by the candidate and others, **the candidate is required to make an explicit statement in the thesis as to who contributed to such work and to what extent.** Supervisors must attest to the accuracy of such statements at the doctoral oral defence. Since the task of the examiners is made more difficult in these cases, it is in the candidates interest to make perfectly clear the responsibilities of all the authors of the co-authored papers.

Thus, chapters two through four of this thesis have an abstract, introduction, methods, results, discussion and references; chapter five, a theoretical paper, omits methods and results sections. Also, as required by the Guidelines, there is a

common abstract, a general introduction (chapter one) and a general discussion (chapter six) which includes claims to originality and suggestions for further research.

The submitted manuscripts are as follows:

- Chapter 2.Fuks A, **Weijer C**, Freedman B, Shapiro S, Skrutkowska M, Riaz A. A study in contrasts: eligibility criteria in a twenty-year sample of NSABP and POG clinical trials. (Submitted for publication).
- Chapter 3. **Weijer C**, Freedman B, Shapiro S, Fuks A, Skrutkowska M, Sigurjonsdottir M. Measuring the interpretation of criteria for clinical trial eligibility: a survey of 365 oncology investigators. (Submitted for publication).
- Chapter 4. Shapiro S, **Weijer C**, Freedman B. Reporting the study populations of clinical trials: clear transmission or static on the line? (Submitted for publication).
- Chapter 5. **Weijer C**. Selecting subjects for participation in clinical research: one sphere of justice. (Submitted for publication).

The candidate was responsible for the following work in the above papers:

- Chapter 2: the candidate participated in discussions regarding the planning and conduct of the research, designed the questionnaire to evaluate inter-rater reliability, assisted in collecting the data, performed the data analysis (under the supervision of Prof. Shapiro), and wrote the first draft of the paper.

● Chapter 3: the candidate participated in planning the research, assisted in designing the questionnaire, conducted the questionnaire mailings, data entry and data editing, performed the data analysis (under the supervision of Prof. Shapiro) and wrote a substantial portion of the first draft of the paper (introduction, methods, results, and part of the discussion, plus all figures and tables).

● Chapter 4: the candidate participated in planning sessions for the project, assisted in data collection, assisted in data analysis, and wrote the methods and results sections and prepared all figures and tables for the first draft of the paper.

● Chapter 5: the candidate is the sole author of the paper.

The candidate's work is supported by a fellowship from the Medical Research Council of Canada. Throughout the preparation of the candidate's doctoral thesis, helpful advice and comments on earlier versions of papers was received from his colleagues in the Clinical Trials Research Group, McGill University: Benjamin Freedman, Ph.D., Abraham Fuks, M.D., C.M., F.R.C.P.(C), Stanley Shapiro, Ph.D., Kathleen Cranley Glass, D.C.L., Karen Lebacqz, Ph.D., Trudo Lemmens, LL.L. and Myriam Skrutkowska, B.Sc.N..The candidate would like to express his sincere gratitude to his colleagues for their encouragement and support during the preparation of this thesis.

He would also like to thank Anthony Belardo, M.A.(*cand.*) and Robert Crouch, M.A. who provided editorial assistance with the thesis manuscript. The

candidate is particularly grateful to his supervisors, Professors Benjamin Freedman and Abraham Fuks, for their intellectual and moral guidance. To them he owes a debt which only a lifetime of scholarly work could repay.

Wisdom is as good as an inheritance,

a real profit for mankind;

for wisdom like wealth is a defence,

but knowledge does more good than money,

it safeguards a man's life.

...

Who is like a wise man?

Who can explain things?

Man's wisdom lights his face up,

it transfigures even a rough countenance.

Ecclesiastes 7:11,12 and 8:1

Table of Contents

Chapter 1:

Introduction:

The *Belmont Report* framework, evolving ethical issues in the selection of subjects for clinical research, and questions for study.

Introduction

In 1994, the U.S. National Institutes of Health (NIH) released a policy, entitled *NIH Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research*, mandating the inclusion of women and members of minority groups in NIH-funded research studies.[1] The policy additionally requires that phase III clinical trials (large clinical trials aimed at changing medical practice) examine potential differences in intervention effect between genders or among racial groups, and that investigators define a program for enrolling and retaining in the study women and members of minority groups. While these ends may not be practical in some cases (for example, when the disease in question is specific to one gender or racial group), the policy does not allow for exemptions based on the cost that such additional procedures may incur. Institutional Review Boards (IRBs), committees that review such research for ethical acceptability, are one of the parties charged with the task of ensuring that investigators abide by these new requirements.

Research policies prior to the *NIH Guidelines* have addressed ethical issues related to the selection of subjects for clinical research. As we shall see, these policies focused on the inappropriate inclusion of so-called vulnerable groups or the wrongful exclusion of groups of individuals who may benefit from research participation. The innovation of the *NIH Guidelines* is the recognition that

the wide-spread applicability of knowledge generated from phase III clinical trials is a scientific and ethical imperative. From a scientific perspective, research that aims to change clinical practice ought to be based on study populations that mirror patients in clinical practice. From an ethical viewpoint, the knowledge arising from clinical research is a social good that ought to be distributed equitably. The systematic exclusion of segments of the patient population from research studies is, therefore, problematic for both ethical and scientific reasons. The *NIH Guidelines* represent an attempt to address this problem:

> Since a primary aim of research is to provide scientific evidence leading to a change in health policy or a standard of care, it is imperative to determine whether the intervention or therapy being studied affects women or men or members of minority groups and their subpopulations differently. To this end, the guidelines published here are intended to ensure that all future NIH-supported biomedical and behavioral research involving human subjects will be carried out in a manner sufficient to elicit information about individuals of both genders and the diverse racial and ethnic groups and, in the case of clinical trials, to examine differential effects on such groups.[2]

Scientific and ethical concerns in research are often viewed -- rightly or wrongly -- as non-overlapping. Scientists may see requirements such as informed consent as purely ethical matters, and without scientific implication; ethicists may view aspects of the study protocol, for example sample-size calculations, as purely scientific and without ethical implication. Selection procedures for participation in clinical research, often operationalized within study protocols as eligibility criteria, have *clear ethical and scientific implications*. Such selection procedures, therefore, represent a fascinating opportunity to examine one area of overlap between science and ethics.

The purpose of this initial chapter is three-fold. First, to lay out the predominant framework for the analysis of ethical problems in human experimentation. Second, to outline briefly the evolving ethical issues in the selection of subjects for clinical research. Third, and finally, to present the main questions addressed in this thesis.

The *Belmont Report*

The U.S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (hereafter, the "National Commission") was created when the National Research Act was signed into law on July 12, 1974.[3] The National Commission was charged, *inter alia*, with defining a set of ethical

principles which could serve to guide the conduct of research involving human subjects. The objective was to "provide an analytical framework that will guide the resolution of ethical problems arising from research involving human subjects."[4] In their final publication, the *Belmont Report*, the members of the National Commission lay out three such principles, respect for persons, beneficence, and justice.

Respect for persons. The principle of respect for persons requires that individual autonomy be acknowledged and that persons with diminished autonomy be protected. An autonomous person is "an individual capable of deliberation about personal goals and of acting under the direction of such deliberation."[5] The choices of such persons should be respected and ought not be interfered with unless there is a clear risk of harm to others. Not all persons, however, are capable of autonomous choice. Some, such as children or the mentally infirm, may lack the capacity for "deliberation about personal goals;" others, perhaps including prisoners, may be in circumstances that restrict their liberty so severely as to bring into question their capacity for free choice. Such persons, according to the National Commission, are entitled to added protections, the degree of additional protection depending on the probabilities of harm and benefit presented by an individual research study.

The principle of respect for persons finds expression most obviously in the

requirement for informed consent to research participation. In order for consent to participate in research to be valid, consent must be based on adequate information, the information must be understood sufficiently and consent must be given freely. How much must research subjects be told? The members of the National Commission propose the standard of the "reasonable volunteer;" in other words, potential subjects must be told information that a reasonable person in that situation would need to know to make an informed decision. This information will likely include: the purpose of the research, procedures involved, potential benefits and harms, alternatives to study participation, and the fact that subjects have the right to ask questions and to withdraw from the study at any time. Researchers have an obligation to present the information in a comprehensible manner and to make efforts to ascertain that subjects have understood the information provided. Finally, potential research subjects must neither be coerced (i.e., threatened or bullied into participation) nor subjected to undue influence (i.e., offered an excessive reward for participation); the decision whether to participate must be made freely.

Beneficence. The principle of beneficence requires that persons not only be protected from harm, but also that steps be taken to ensure their well being. The principle is operationalized by two complementary rules: first, do not harm, and, second, maximize possible benefits and minimize possible harms. In the context of

research, investigators (and Institutional Review Boards) have an obligation to maximize potential benefits and reduce risks associated with individual research projects.

Risk refers to both the probability and magnitude of potential harm and is properly compared with potential benefits. Potential harms and benefits may accrue to the individual research subject, their families, and to society in general; a thorough analysis of risks and potential benefits requires that all of these be examined. The National Commission recommended that ethically acceptable research should, at a minimum, reflect the following requirements: the treatment of research subjects should never be inhumane; risks should be minimized (in accord with the exigencies of science); the assessment of risk should be particularly scrupulous when the study population involves persons with diminished autonomy; studies should justify the inclusion of vulnerable groups; and relevant risks and potential benefits should be fully and accurately disclosed in the consent process.

Justice. Justice, as conceived by the members of the National Commission, refers to the fair distribution of goods; in the context of research, it refers to the equitable distribution of the risks and potential benefits of research participation. As the principle of respect for persons is the foundation for requirements for informed consent, so too the principle of justice provides the underpinnings for the

obligation that research subjects be selected fairly.

IRBs have an obligation to scrutinize the selection of subjects for clinical research to ensure that fair procedures are implemented. On the level of the individual, researchers ought neither select patients who they like for potentially beneficial research nor choose "undesirable" patients for potentially harmful research. On a societal level, IRBs ought to ensure, for example, that classes of persons are not "being systematically selected simply because of their easy availability, their compromised position, or their manipulability, rather than for reasons directly related to the problem being studied."[6] Social justice may require that some classes of subjects be selected before others for research participation; for example, *ceteris paribus*, autonomous adults should be enrolled in research before persons of diminished autonomy. As we shall see, the nature of justice-related concerns in research has changed over the last decades.

Evolving ethical issues in selection of subjects for clinical research

Although we have characterized justice in the context of research as the "equitable distribution of the burdens and benefits of research participation," I have argued elsewhere that the emphasis and scope of justice-related issues in clinical research have evolved over time.[7] (The following is a summary of these earlier findings.) Early concerns in research ethics were fueled by the revelation of

research scandal and focused on protecting so-called vulnerable groups from the potential burdens associated with research participation. As research participation came to be seen as a potential benefit, because either clinical care within studies was thought to be superior or experimental treatments were available only in studies, ethicists argued that unnecessary barriers to research participation ought to be removed. Most recently, concern has been expressed that the systematic exclusion of certain groups from research has led to insufficient knowledge regarding the optimal treatment of persons from such groups. How, then, did each of these concerns develop? And what were (and are) the implications for the regulation and conduct of clinical research?

Protecting the vulnerable from harm. From the end of the Second World War until the early 1970s, a number of research scandals highlighted the inappropriate inclusion of vulnerable groups in research. As Freedman has pointed out, the ethical violations of these scandals were multi-dimensional.[8] In some cases, informed consent was not obtained at all, in others subjects were informed incompletely or deceived. Many of the studies presented a poor balance of benefits and harms, ranging from research that deprived subjects of needed treatments, to that which knowingly and predictably harmed subjects. Subjects for such experimentation were drawn from vulnerable or "undesirable" classes of persons, including the mentally infirm or demented, political prisoners, racial minorities,

the poor and the under-educated. Taken as a whole, this unethical research generated the belief that participation in research was a risky venture, one from which persons would wish to be protected.

Perhaps the best known example of unethical research is the heinous human experimentation carried out in Nazi Germany during World War II. German physicians and scientists subjected Jews, Russians, Gypsies, political prisoners, homosexuals[9] and others to a wide range of research.[10] The efforts of the Nazi state to eliminate non-Aryan people (the policy of "racial hygiene") led to experiments examining sterilization techniques and methods of mass murder. Other research studies were motivated by the exigencies of war; for example, the hypothermia experiments at Dachau were sparked by high losses of Axis aviators shot down in the North Sea.[11] In the Dachau experiments, research subjects were immersed in tanks of ice water and either observed until death or a variety of revival techniques were tested. Approximately 25% of the research subjects died as a direct result of their participation in the hypothermia experiments.

The first widely-publicized research scandal in the United States involved three physicians at the Jewish Chronic Disease Hospital in Brooklyn, New York.[12] As a part of a larger research project examining the immunology of cancer, twenty-two long-term care patients in the hospital were injected with suspensions containing live cancer cells. Although all of the study participants were chronically

ill, none had cancer. Problematically, participants in the study were not informed of the fact that the injections contained cancerous cells. Furthermore, at the time that the scandal broke, it was alleged that many of the participants were incapable of giving valid informed consent. Fortunately, as the investigators had hypothesized (but not known *ab initio*), none of the patients developed cancer as a result of the injections.

The Tuskegee syphilis experiment remains one of the most widely-known examples of unethical research in the United States.[13] Perhaps the longest running research project funded by the U.S. Public Health Service (1932-1972), the Tuskegee syphilis study examined the course of untreated syphilis in four hundred Afro-American men in rural Alabama. Study participants were misinformed and told that invasive tests, such as spinal taps, done solely for research were "treatments." Furthermore, when penicillin, a safe and highly-effective treatment for the disease, became available after World War II it was withheld from study subjects. It is estimated that twenty percent of the study participants died prematurely.

Against this backdrop of scandal and deceit it is not surprising that research participation was thought to be a risky business. Early writers on the ethics of research were preoccupied with protecting potentially vulnerable groups from the burdens of research participation. Indeed, the members of the National

Commission required that the involvement of certain groups, including "hospitalized patients, or other institutionalized persons, or disproportionate numbers of racial or ethnic minorities or persons of low socioeconomic status should be justified."[14] The notion that certain classes of subjects need to be protected survives into current Department of Health and Human Services regulations:

> (3)[the IRB must determine that] Selection of subjects is equitable. In making this assessment the IRB should take into account the purposes of the research and the setting in which the research will be conducted and should be particularly cognizant of the special problems of research involving vulnerable populations, such as children, prisoners, pregnant women, mentally disabled persons, or economically or educationally disadvantaged persons
>
> ...
>
> (7)(b)When some or all of the subjects are likely to be vulnerable to coercion or undue influence, such as children, prisoners, pregnant women, mentally disabled persons, or economically or educationally disadvantaged persons, additional safeguards have been included in the study to protect the rights and welfare of these subjects.[15]

Study participation as a benefit. In the 1980s the public perception of

research participation changed substantially. Levine observes that

> what was once seen as threatening — a burden from which people would wish to be protected — is now seen as a benefit. People are clamoring for access to clinical trials and to experimental drugs. People are demanding that they, and others who are like them, are owed such as a matter of justice.[16]

Several factors were responsible for this shift. Between 1976 and 1982, a number of studies were published which for the first time examined the risk associated with research participation.[17] The studies found that the level of risk presented by research participation was relatively small. Research studies with therapeutic interventions, particularly those examining the treatment of cancer, seemed to present more risk than studies without such interventions, but the risks in such studies did not differ, at least in kind, from those present in clinical practice.[18] The advent of HIV/AIDS in the early 1980s was, however, the major catalyst for the change in perception. In the early years of the epidemic, no effective treatments were available and experimental treatments seemed to many to offer the best hope for survival.

As a result of these changes, many of the groups excluded from research participation for protection, were now seen to be disadvantaged or even harmed by being denied access to experimental treatments available only in clinical trials. As

a result, the ethical issue of the day changed from the equitable distribution of the burdens of research participation to the *fair distribution of the benefits of research participation*. For example, Carol Levine asks

> How can groups of prospective subjects traditionally excluded from clinical trials because of their physical or social vulnerability (women of child bearing age, infants, prisoners, intravenous drug users, prostitutes) be given access to clinical trials that *may*, perhaps, prove of benefit to them?[19]

Ethicists and regulators responded to this question by arguing (and requiring) that groups of subjects not be excluded from research participation without good reason. In the context of HIV/AIDS research, trial designers have been advised that

> Criteria for inclusion in phase II and III clinical trials should be based on a presumption that all groups affected by the research are eligible, regardless of gender, social or economic status, use of illicit drugs, or stage of illness unless the study is particularly designed to look at a particular stage of illness.[20]

Such requirements have not yet made their way into Department of Health and Human Services regulations. However, the Office for Protection from Research Risk's *Institutional Review Board Guidebook* does ask IRBs to consider the

following questions when reviewing subject selection procedures in research proposals:

> To the extent that benefits to the subject are anticipated, are they fairly distributed? Do other groups of potential subjects have a greater need to receive any of the anticipated benefit?
>
> Has the selection process *overprotected* potential subjects who are considered vulnerable (e.g., children, cognitively impaired, economically or educationally disadvantaged persons, patients of researchers, seriously ill persons) so that they are denied opportunities to participate in research?[21]

The importance of widely-applicable research results. Recently, a new justice issue has been added: the results of research ought to be applicable to the wide range of affected persons in society. The concern was first raised in the context of HIV/AIDS: women, children and other groups were excluded from early treatment studies and, thus, little was known about how best to treat these groups of HIV/AIDS sufferers.[22] In the early 1990s the issue expanded in scope dramatically. In an influential article in 1992, Dresser claimed that "the failure to include women [and members of racial and ethnic minority groups] in research populations is ubiquitous."[23] The exclusion of such groups from research, said

Dresser, makes it inappropriate to conclude that new treatments are safe and effective in groups not included in research studies.

Politicians responded to the ensuing public outcry with a number of important measures. First, the U.S. Food and Drug Administration (FDA) removed existing barriers to the participation of women of reproductive potential in early-stage clinical research testing the safety and efficacy of new drugs.[24] In the new guidelines, the FDA acknowledges that "[t]he patients included in clinical studies should, in general, reflect the population that will receive the drug when it is marketed."[25] Second, as discussed in the opening section of this chapter, in 1994, the U.S. NIH released guidelines requiring the inclusion of women and members of racial and ethnic minorities in all NIH-funded research.

Questions for study

Given the wide spread recent interest in the ethical and scientific implications of selection procedures for clinical research participation, a systematic examination of criteria for clinical trial eligibility is both timely and important. As outlined below, I will examine selection procedures for research participation at a number of points in the genesis and dissemination of medical knowledge: the clinical trial protocol, the interpretation of criteria by clinical investigators, and the communication of study results. Also, a variety of study

methods will be employed, including both empirical and theoretical (philosophical) approaches.

In their study of a set of concurrent breast cancer clinical trials, Begg and Engstrom report that oncology studies often contain numerous eligibility criteria and that many of these criteria are seemingly arbitrary.[26] The number of restrictive criteria in trials is so great, they conclude, as to bring into question the wide spread applicability (generalizability) of such narrowly focused studies. Building on Begg and Engstrom's important work, colleagues and I sought to answer a number of related questions:

- How have numbers of criteria in comparable trials changed over time?
- What is the nature of criteria that have been added or dropped?
- How do these changes affect the generalizability of study results?

In order to address the questions, we undertook an empirical study of eligibility criteria found in cancer clinical trials funded over a twenty-year period by two important cooperative groups: the Pediatric Oncology Group (POG) and the National Surgical Adjuvant Breast and Bowel Project (NSABP). The results of this study are presented in chapter 2.

In our study of POG and NSABP eligibility criteria, my colleagues and I were struck by the fact that some eligibility criteria were phrased subjectively (e.g., "patients must have a life expectancy of at least 10 years, excluding their

diagnosis of cancer, to be enrolled in the study") whereas others were phrased objectively (e.g., "to be included in the study the patient must have a white blood cell count (WBC) greater than or equal to 4,000/ mm^3 and a platelet count greater than or equal to 100,000/ mm^3"). Hypothesizing that subjective criteria could be a source of variability in decisions by investigators to enroll patients and investigator uncertainty, we sought to answer two questions:

- Are subjective eligibility criteria associated with more variable enrollment decisions?
- Are investigators less certain of their decisions when using such criteria?

In order to answer these questions, we surveyed 365 oncologist-investigators from the United States and Canada. The results of this study are presented in chapter 3.

Eligibility criteria are not merely of importance in the planning and conduct of research. Eligibility criteria must be accurately and completely reported in communications of study results for a couple of reasons. First, other investigators can only replicate a study if eligibility criteria are fully reported. Second, and more important, clinicians in practice need access to the full list of eligibility criteria in order to make an accurate assessment as to which patients in their clinical practice the results of a particular study apply. Colleagues and I, therefore, undertook to answer the following questions:

- Are eligibility criteria accurately reported in communications regarding the

results of clinical trials?

- If reporting is not complete, what is the nature of information loss (i.e., what types of criteria are not reported)?

In order to examine these questions, we studied the reporting of eligibility criteria in study protocol, methods paper, journal article, and Clinical Alert issued by the U.S. National Institutes of Health of eight important clinical trials. The results of this study are presented in chapter 4.

Finally, if recent regulatory requirements regarding selection procedures for clinical research participation are to be applied optimally, we need a comprehensive theoretical understanding of just and unjust eligibility criteria. One possible starting point for this philosophical problem is Michael Walzer's view of the complex egalitarian society presented in *Spheres of Justice*.[27] I ask:

- What are the implications of Walzer's political philosophy for the just selection of subjects for research participation?

The fifth chapter contains the results of this philosophical analysis.

The sixth and final chapter of the doctoral thesis discusses the overlap of scientific and ethical concerns in clinical research, reviews the methodology presented in this thesis and suggests how it may be applied to other areas of scientific and ethical concern in medical research.

References

1.Department of Health and Human Services, National Institutes of Health. NIH guidelines on the inclusion of women and minorities as subjects in clinical research. *Federal Register* 1994; 59: 14508-14513.

2.*ibid.*

3.National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: ethical principles and guidelines for the protection of human subjects of research. *OPRR Reports*; April 18, 1979: 1-8.

4.National Commission, *op. cit.*, p. 3.

5.National Commission, *op. cit.*, p. 4.

6.National Commission, *op. cit.*, p. 5.

7.Weijer C. Evolving ethical issues in selection of subjects for clinical research. *Cambridge Quarterly of Healthcare Ethics* 1996; 5: 334-345. This paper was based on work presented in my Master's Thesis: Weijer C. Characterizing the population in clinical trials: barriers, comparability, and implications for review. Master's Thesis. McGill University, Montreal, 1995.

8.Freedman B. Unethical research. In: W.T. Reich (ed.). *Encyclopedia of Bioethics*. New York: Simon and Schuster MacMillan, 1995: pp. 2258-2261.

9.Levay S. *Queer Science: The Use and Abuse of Research into Homosexuality.* Cambridge, M.A.: MIT Press, 1996.

10.Caplan, AL. (ed.). *When Medicine Went Mad: Bioethics and the Holocaust.*

Humana Press. Totawa, N.J., 1992.

11.Berger, RL. Nazi science – the Dachau hypothermia experiments. *New England Journal of Medicine* 1990; 322: 1435-1440.

12.Katz J. *Experimentation with Human Beings*. New York: Russell Sage Foundation, 1972: pp. 9-65.

13.Jones, JH. *Bad Blood: The Tuskegee Syphilis Experiment*. New York: The Free Press, 1993.

14.National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *Report and Recommendations: Institutional Review Boards* [DHEW Publication (OS) 78-0009]. Washington, D.C.: Department of Health, Education, and Welfare, 1978: p. 22.

15.Department of Health and Human Services. Protection of Human Subjects. Title 45, *Code of Federal Regulation*. Part 46: Revised. 1991; June 18: 46.111(a).

16.Levine RJ. The impact of HIV infection on society's perception of clinical trials. *Kennedy Institute of Ethics Journal* 1994; 4: 93-98.

17.Cardon PV, Dommel FW, Truble RR. Injuries to research subjects: a survey of investigators. *New England Journal of Medicine* 1976; 295: 650-654; Zarafonetis CJD, Riley PA, Willis PW, et al.. Clinically significant adverse events in a phase one testing program. *Clinical Pharmacology and Therapeutics* 1978; 24: 127-132; McCann DJ, Pettit JR. A report on adverse effects insurance for human subjects. In: President's Commission for the Study of Ethics Problems in Medicine and

Biomedical and Behavioral Research. *Compensating for Research Injuries: The Ethical and Legal Implications of Programs to Redress Injuries Caused by Biomedical and Behavioral Research* (Appendices). Washington, D.C.: U.S. Government Printing Office, 1982; Arnold JD. Incidence of injury during clinical pharmacology research and indemnification of injured research subjects at the Quincy Research Center. In: President's Commission for the Study of Ethics Problems in Medicine and Biomedical and Behavioral Research. *Compensating for Research Injuries: The Ethical and Legal Implications of Programs to Redress Injuries Caused by Biomedical and Behavioral Research* (Appendicies). Washington, D.C.: U.S. Government Printing Office, 1982.

18.Arnold, *op. cit.*; Cardon, *op. cit.*.

19.Levine C. Has AIDS changed the ethics of human subjects research? *Journal of Law, Medicine, and Health Care* 1988; 16: 167-173.

20.Levine C, Dubler CC, Levine RJ. Building a new consensus: ethical principles and policies for clinical research in HIV/AIDS. *IRB: A Review of Human Subjects Research* 1991; 13(1-2): 1-17.

21.Office for the Protection from Research Risk. *Protecting Human Research Subjects: Institutional Review Board Guidebook.* Washington, D.C.: U.S. Government Printing Office, 1993: 3.26.

22.Macklin R, Friedland G. AIDS research: the ethics of clinical trials. *Journal of*

Law, Medicine and Health Care 1986; 14: 273-280.

23.Dresser R. Wanted single, white male for medical research. *Hastings Center Report* 1992; 22: 24-29.

24.Food and Drug Administration. Guideline for the study and evaluation of gender differences in the clinical evaluation of drugs. *Federal Register* 1993; 58: 39406-39416.

25.*ibid.*

26.Begg CB, Engstrom PF. Eligibility and extrapolation in cancer clinical trials. *Journal of Clinical Oncology* 1987; 5: 962-968.

27.Walzer M. *Spheres of Justice: A Defense of Pluralism and Equality*. New York: Basic Books, 1983.

Chapter 2:

A study in contrasts:

eligibility criteria in a twenty-year sample of NSABP and POG clinical trials.

Abstract

Background. While a number of barriers to enrollment in clinical trials have been identified in the literature, eligibility criteria are the most important impediment to accrual. We sought to study changes in criteria in two diachronous samples of clinical trials.

Methods. Clinical trials from two cooperative groups, NSABP (n=11) and POG (n=7), covering a twenty-year time span (1972-1992) were included in the study. After duplications were eliminated, the criteria in each protocol were enumerated and classified according to a novel schema.

Results. The NSABP trials contained more criteria (36.7 [mean] ± 1.70 [SE]) than the POG studies (9.4 ± 1.38, p=0.0001). NSABP studies added precision criteria (criteria that attempt to make the study population more homogeneous) at a faster rate than POG studies. Furthermore, NSABP protocols contained criteria that excluded patients thought to be vulnerable to toxicity whereas the POG trials did not.

Conclusions. The NSABP studies typify an explanatory approach to clinical-trial design, whereas the POG trials typify a pragmatic approach. The contrast between the two groups suggests that large numbers of eligibility criteria may not be necessary for good quality studies. We recommend that: (1) the distinction between inclusion and exclusion criteria be abandoned; (2) eligibility criteria

should be explicitly justified; (3) the need for each criterion ought to be assessed when a new trial in a series is planned; (4) criteria in phase III clinical trials should restrict the eligible patient population as little as possible; and (5) research is needed to define the impact of criteria on generalizability. (250 words)

Key words. Clinical trials, eligibility determination, patient participation, neoplasms/ treatment.

Background

Clinical trials are the keystone in the development and evaluation of new treatments in oncology. Despite the pivotal role of clinical trials, only a small proportion of cancer patients are actually treated in trials. Friedman and colleagues report that only 1.6% of U.S. cancer patients are enrolled in NCI-funded phase II and III clinical trials.[1] Low accrual rates have been identified repeatedly as a critical problem affecting U.S. cancer trials.[2,3]

Why are so few patients enrolled in clinical trials? Many cancer patients never have the opportunity to enroll in a research study either because they are treated in a hospital that doesn't participate in multi-center clinical trials or because no trials are available for their type and stage of disease. The NCI's Community Clinical Oncology Program (CCOP) aims towards -- and has been successful in -- increasing the involvement of community hospitals in cancer clinical trials.[4,5,6,7] But even within institutions that actively participate in clinical trials, only a minority of cancer patients for whom a study is available are treated in studies.[8]

Much of what has been written on barriers to clinical trial enrollment in oncology has focussed on physician[9,10,11,12,13,14] and patient[15,16,17,18] factors influencing accrual to research studies. In fact, however, criteria for clinical trial eligibility are the largest barrier to trial accrual. McCusker and colleagues describe

the enrollment experience of a cohort of 454 patients in a single medical oncology clinic: of the 342 patients for whom a trial was available for their type and stage of disease, 43% were ineligible, 4% were not enrolled due to physician refusal, 2% refused to give consent, 1% were not enrolled for "other" reasons and only 25% were actually enrolled in a research study.[19] Kotwall and colleagues describe the enrollment experience of a cohort of 592 women with stage I or II breast cancer (a protocol was available for all of the women): 46% were ineligible, 5% were not enrolled due to physician refusal, 25% refused to provide consent, 6% were not enrolled for "other" reasons and only 18% were actually enrolled in a research study.[20] Other studies of accrual of oncology patients to trials have reported similar results.[21,22]

The exclusion of patients from cancer research is an important issue for a number of reasons. Patients who are barred from study participation may be deprived of benefits associated with trial treatment.[23] A number of studies have observed that cancer patients treated in clinical trials have better outcomes than those who are treated outside of trials.[24,25,26] Weijer and colleagues examined treatment differences between women with early stage breast cancer treated within and outside clinical research studies at a single institution.[27] Their findings suggest a potential mechanism for the observed survival advantage for trial participants: when age and stage of disease were controlled for in the analysis, women in trials

received higher doses of chemotherapy and more frequent blood tests than other patients.

Another problem with exclusions is that trials that prevent substantial proportions of patients from participating may yield results that are not widely applicable (i.e., not generalizable) to the broader patient population of interest. Optimally, clinical trials should mirror the patient population in clinical practice; such trials will maximize accrual rates and foster the widespread applicability of results essential to maximize a trial's impact on medical practice.[28] Advocates of more restrictive trials assert that narrow trials are more efficient because patient heterogeneity (i.e., variance) is reduced.[29] In oncology, however, we know too little about prognostic factors to be able to define a truly homogenous population of patients.[30] Furthermore, both Buyse and George have independently argued that including patients of differing prognoses can allow a trial to accrue patients more quickly, and thus answer the question of interest more efficiently (in terms of time required to complete a trial).[31,32]

How many eligibility criteria do typical phase III cancer trials have? Are all the criteria necessary? In an attempt to answer these questions, Begg and Engstrom studied a synchronous sample of trials investigating chemotherapy in the treatment of stage II, node positive breast cancer.[33] They found that each of the studies contained a large number of criteria defining the eligible patient population (the

average number of criteria per study was 23). They also found that, among these otherwise comparable trials, substantial variation existed in eligibility criteria. Begg and Engstrom conclude that "[t]he rationale for these exclusions is not clearly understood in many cases and may to some extent be due to the unchallenged perpetration of conventions that are more applicable to laboratory experiments in which the experimenter is able to exercise much more control of the conduct of the study." The magnitude and nature of the exclusion criteria were such as to "cast doubt on the generalizability of the results from the clinical trials program."

Crucial questions remain regarding criteria for trial eligibility: How have the numbers of criteria in comparable trials changed over time? What is the nature of the criteria that have been added or dropped? How do these changes affect the generalizability of study results? In order to begin to address these problems, we undertook a study of two diachronous samples: clinical trials of the National Surgical Adjuvant Breast and Bowel Program (NSABP) and the Pediatric Oncology Group (POG).

Methods

Clinical trials from the NSABP and POG were selected because they each constitute a single cooperative group that has focussed a series of clinical trials on

a single disease: in the case of the NSABP, early-stage breast cancer; for POG, acute lymphocytic leukemia. Furthermore, both cooperative groups are the source of distinguished and influential research results and recommendations. Finally, both groups of studies cover a similar time span: 1972 to 1992. Trials were chosen up to the temporary suspension of NSABP studies. While all of the POG acute lymphocytic leukemia studies were included, we selected a subset of the NSABP trials for study. Of the 22 NSABP studies (B-04 to B-25; protocols for B-01 to B-03 were no longer available either locally or from the NSABP central office), we included, on the basis of an a priori decision, only studies of the treatment with chemotherapy of stage II, node positive disease (the same type and stage of disease as the trials studied by Begg and Engstrom).

The inclusion and exclusion criteria from each of the clinical trials were extracted from the protocols. It was immediately apparent that substantial duplications between inclusion and exclusion criteria were present and these were eliminated. (E.g. -- from NSABP B-15 -- inclusion criterion: "Patients must have a life expectancy of at least 10 years excluding their diagnosis of cancer;" exclusion criterion: "Patients who have a life expectancy of fewer than 10 years, excluding their diagnosis of cancer.") The criteria in each study were then counted by at least two of the study authors.

A schema for classifying eligibility criteria was developed by our group.

The schema divides eligibility criteria into categories that describe the purpose of each criterion. By a process of group discussion and negotiation, a schema with five categories (plus a catch-all "other" category) was arrived at. The five main categories with examples from NSABP B-15 are as follows:

- Definition of disease -- eligibility criteria that define the medical condition of interest and represent factors that would be taken into account in clinical practice. For example: "On clinical examination, the tumor was 4 cm or less in its greatest dimension."
- Precision -- eligibility criteria concerned with the scientific validity of the study. These criteria attempt to diminish variability in the study by either making the patient population more homogenous or reducing measurement error. Precision criteria involve factors that would not ordinarily be taken into account in clinical practice. For example: "Patients with a previous malignancy, regardless of site [are excluded]."
- Safety -- eligibility criteria that exclude persons thought to be unduly vulnerable to treatment in general or one of the study treatments in particular. For example: "The post-operative WBC $\geq$ 4,000/ mm^3 and platelet count $\geq$ 100,000/ mm^3."
- Ethical and legal -- eligibility criteria that are required in order to ensure conformity with Department of Health and Human Services regulations governing the conduct of human experimentation. For example: "The patient consents to be

in the study."

● Administrative -- eligibility criteria which attempt to ensure the smooth functioning of the study. Administrative criteria include measures aimed at ensuring compliance with treatment and follow-up fall into this category. For example: "Patient is accessible geographically for follow-up."

Using 42 criteria from one of the NSABP studies (B-22), the inter-rater reliability of the schema was assessed (appendix 2) with five clinical investigators and two of the study authors (AF and CW).[34] The schema proved to have a very good inter-rater agreement, κ=0.77.

Criteria from each of the studies were classified into one of the schema categories by at least two of the study authors. Unfortunately, none of the studies provided an explicit rationale for each eligibility criterion and disagreement over classification of individual criteria was resolved by discussion and consensus.

Non-parametric correlations (Kendall's tau-b) and p-values were calculated for number of eligibility criteria (and date of first accrual) for the two samples using PROC CORR in SAS. Linear regression models were calculated for each of the schema categories (with date of first accrual as the independent variable) using PROC REG in SAS. 95% confidence intervals were calculated for linear regression coefficients.

Results

Of the twenty-two NSABP breast cancer trials between 1972 and 1992, 11 examined the role of chemotherapy in the treatment of stage II, node positive breast cancer (NSABP protocols B-05, B-07, B-08, B-09, B-10, B-11, B-12, B-15, B-16, B-22 and B-25). The number of criteria for all studies is shown in figure 1. The number of eligibility criteria in the larger set of 22 studies increased from 21 to 44 over the twenty year period (coefficient of correlation, *tau*=0.54, p=0.0006). In the subset of 11 trials (hereafter "NSABP subset"), the number of criteria increased from 26 to 44 over a similar time period (*tau*=0.93, p=0.0001). Despite the obvious treatment complexity in the POG studies (ALinC-10 through ALinC-15b), the number of criteria per study (9.4 ± 1.38 [mean ± SE]) was significantly less than NSABP-subset trials (36.7 ± 1.70, p=0.0001). Over the twenty-year time span, the number of criteria in the POG studies increased from 6 to 12 (*tau*=0.71, p=0.04).

The sorting of eligibility criteria for trials in the two groups according to the classification schema is shown in table 1. In the NSABP subset, the majority of the criteria fell into the definition of disease (41%) and precision categories (36%). A minority of the criteria fell into the categories of safety (11%), ethical and legal (9%) and administrative criteria (3%). Significant increases in the number of criteria over time were seen only in the definition of disease (*tau*=0.88, p=0.0004)

and precision categories (*tau*=0.88, p=0.0003). The rate of increase for each of the categories is shown in table 1. On average, definition of disease criteria increased by 0.41 criteria per year and precision criteria increased by slightly more than this, 0.44 criteria per year.

In the POG studies, most of the criteria fell into one of three categories: definition of disease (38%), precision (27%) and administrative criteria (24%). A few criteria were categorized as ethical and legal (11%) and no safety criteria were present in any of the studies. A clearly significant increase over the time period was seen only in definition of disease criteria (*tau*=0.87, p=0.01). Marginally significant increases were seen in administrative (*tau*=0.71, p=0.04) and precision categories (*tau*=0.60, p=0.08). The rate of increase for definition of disease criteria was 0.19 criteria per year. In table 1 and table 2, note that the 95% confidence intervals around the estimates for definition of disease and precision criteria in the NSABP and POG studies do not overlap. We may conclude, therefore, that the NSABP studies added criteria to these two categories at a greater rate than the POG studies did.

In an attempt to elucidate a mechanism for the observed increase in criteria over time, we undertook a closer study of the criteria in the NSABP subset. We discerned 53 distinct criteria in the 11 clinical trials (figure 2). Twenty-two criteria were present in all 11 clinical trials. Of the 26 criteria added to trials after B-05 but

before B-25 (i.e., criteria that were *added to the series* and *at risk of being retained*), 18 (69%) were present in all subsequent trials. Thus, when an eligibility criterion was added in this series of clinical trials, it was unusual for it to be removed.

Twenty-seven criteria were added to the NSABP subset after protocol B-05. Criteria added to protocol B-07 are representative:

- definition of disease: women with "ipsilateral axillary nodes over 2 cm in greatest diameter" were excluded (criterion #27, figure 2); women with inflammatory carcinoma were excluded (#28);
- precision: "therapy must begin within 2-4 weeks after mastectomy" (#29);
- safety: women must have "evidence of adequate hepatic function" (#30);
- ethical and legal: "patients with psychiatric or addictive disorders" that would preclude informed consent (#31) or prevent them from receiving any of the study treatments (#32) were excluded; and
- no administrative criteria were added.

A complete listing of eligibility criteria in the NSABP-subset studies is given in appendix 1.

Discussion

Until relatively recently, the scope of ethical enquiry into clinical research

was limited to a few discrete areas. Following the final report of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, the Belmont Report, research ethicists have focussed their attention primarily on informed consent and the assessment of risk-benefit ratio in clinical research.[35] Freedman recently reviewed the cumulative index (1979-1990) of the main peer-reviewed journal for research ethics and review, IRB: A Review of Human Subjects Research [unpublished data]. Articles on informed consent (99 articles) and confidentiality (43) represented the largest group, a smaller number of articles was related to risk-benefit assessment (40), and only a few examined ethical issues in the selection of subjects for research (5).

Nonetheless, the scope of ethical concern regarding the selection of subjects for clinical research has expanded over the last thirty years.[36] In the 1970s, ethicists and Institutional Review Boards (IRBs) attempted to ensure that potentially vulnerable subjects (e.g., members of racial minorities, under-educated persons) were not unduly burdened by research participation. With the advent of HIV/AIDS in the 1980s, the concern shifted to ensuring that groups of subjects were not unjustly denied the benefits that might accrue from research participation. Recent developments, including the NIH Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research (hereafter, "NIH Guidelines"), highlight a new domain of ethical interest: the effect of subject selection on the

generalizability of research findings.[37]

Parallelling these evolving concerns, the eligibility criteria in a multi-center clinical trial may come under scrutiny -- and indeed may be changed -- by an IRB for a number of reasons.[38] First, a trial may fail to exclude persons at undue risk from study participation or in need of a particular treatment that they may not receive in the study. Second, a trial may, via an exclusion criterion, bar unjustly a group of patients from the benefits of trial participation. For example, of the 25 cases in which the ethics committee studied by Freedman required changes to eligibility criteria, five involved challenging the exclusion of persons with HIV. Third, eligibility criteria may be so restrictive as to make the accrual of an adequate number of patients to the study unlikely. The NIH Guidelines add another level of scrutiny to eligibility criteria by requiring that the IRB question eligibility criteria that impede directly or indirectly the enrollment of women or minorities to NIH-funded clinical trials.

Our investigation represents, then, an enquiry into this expanding area of ethical (and political) interest. In it, we present a novel avenue of research into eligibility criteria in clinical trials and characterize changes over time in two important groups of studies. We found that eligibility criteria in both the NSABP and POG studies increased substantially over the study period. But an increase in the number of criteria alone is not sufficient to conclude that a problem exists. We

must ask: What is the nature of the increase in these two groups of trials? How is the generalizability of study results affected?

Our classification schema was constructed with the issue of generalizability in mind. For example, the definition of disease category was designed to include criteria that mirror the factors that clinicians incorporate in decisions in medical practice. Increases in definition of disease criteria are, therefore, unlikely to interfere with the clinical applicability of trial results. The validity of our classification is bolstered by the high rate of agreement between the study authors and independent clinician-investigators. Increases in ethical and legal criteria and administrative criteria are also unlikely to impede generalizability of trial results; such criteria, for the most part, involve factors with little biological significance (e.g., "geographical availability"). Changes in precision criteria will, however, likely impact on the applicability of research results: such criteria restrict the patient population by biologically-relevant factors not used in clinical practice (e.g., excluding patients over age 70). So too, the presence of safety criteria may diminish the clinical applicability of trial results if excluded patients would be treated in practice in ways similar to those on the protocol. Given their impact on generalizability, we will focus most of our attention on precision and safety criteria.

The increase in eligibility criteria in both the NSABP and POG clinical

trials was partly accounted for by increases in definition of disease criteria. As we have said, the generalizability of trial results is likely not threatened by changes in such criteria. The greatest increase in criteria in the NSABP trials was accounted for by precision criteria, i.e., criteria that have a direct impact on the applicability of study findings to clinical practice. Precision criteria were added at a greater rate in the NSABP trials than in the POG trials.

Some of the precision criteria are probably not of great clinical import. For example, the criterion requiring that "therapy must begin 2-4 weeks after mastectomy" (NSABP B-07, criterion #27, figure 2) seems innocent enough. Nonetheless, restricting the population in this manner means we have little information on the effect of delaying treatment beyond this narrow window. Furthermore, by restricting the pool of eligible patients by this seemingly trivial criterion (and others like it), the trial designers may diminish the rate at which patients can be accrued to studies. Finally, even if the addition of a single criterion in this category has a relatively minor effect on the generalizability of study results, the addition of a number of such criteria may have a substantial impact *en masse* on clinical applicability.

Other precision criteria are of greater clinical importance. The NSABP trials routinely excluded patients "with previous malignancy" (criterion #12, figure 2), with "concomitant malignancy" (#13), with serious "non-malignant systemic

disease" (#14), who received prior "irradiation ...[or] chemotherapy" (#15), or who received "prior hormonal therapy" (#16). Exclusion of these groups of patients leaves the clinician with no information on the risks and benefits of the investigational treatments in such groups. The clinician may well wonder: Is a woman with breast cancer who has received prior cancer treatment more susceptible to harmful effects from treatment? Do the benefits of the proposed treatment outweigh these risks? Unfortunately, no information can be forthcoming from trials that exclude such groups.

The exclusion of older persons from cancer clinical trials deserves special attention. Despite the fact that the elderly carry the largest share of the burden of cancer, relatively little information regarding cancer therapy in the elderly is available.[39] As a result, the elderly are at risk of being undertreated, a problem that has been connected with the paucity of clinical trials addressing the treatment of cancer in the elderly.[40,41] It is, therefore, a substantial concern that NSABP trials B-08 through B-12 excluded "patients over 70 years of age" (criterion #24, figure 2; protocol B-05 and B-07 excluded "patients over 75 years of age"). While such exclusions may have been motivated in part by a concern that older patients may be more likely to experience toxicity from anticancer treatment, recent data does not seem to bear this out.[42,43] More recent NSABP protocols have replaced this criterion with the criterion "patients must have a life expectancy of at least 10

years excluding their diagnosis of cancer" (#46), but it is uncertain how clinician-investigators in practice interpret this relatively vague requirement. There is no evidence to suggest that larger numbers of older patients are being entered into NSABP breast cancer studies. Indeed, according to 1992 accrual data from the NSABP, the proportion of women over the age of sixty entered into trials actually *dropped* after the criterion's introduction from 24% to 17% ($p<0.0001$). (When NSABP B-15 and B-16 -- studies that excluded women greater than sixty years of age (B-15) and women less than fifty years of age (B-16) -- are dropped from the comparison the magnitude of the drop diminishes: 24% to 20%, $p<0.0002$).

A striking difference between the NSABP and POG trials is the total absence of safety criteria from the POG studies. All of the POG studies involved complicated and relatively intense chemotherapy regimens. Despite this fact, the POG trial designers left the matter of safety to the individual clinician's judgement. The NSABP studies excluded a number of groups of patients with criteria thought to be motivated by safety concerns: patients with abnormal hematologic (criterion #18, figure 2), renal (#19) or hepatic (#30) indices and patients with a history of heart disease (#35). To the extent that safety criteria in a protocol make explicit prudent clinical judgement, they do not impact upon the generalizability of study results. Likely, the criterion relating to Adriamycin toxicity (#35) falls into this category. Other criteria, though, exclude from the

study patients who would be treated in clinical practice. Patients with minimally or moderately abnormal blood, hepatic or renal indices, certainly require treatment. Excluding these groups from trials leaves clinicians with no information on the risks and benefits of treatment in such cases. For some groups of patients, chemotherapy dose-modification may be appropriate, but again the failure to examine such modifications in trials leave clinicians in practice with no information as to how best to proceed.

An extensive list of exclusion criteria may not even protect research subjects from harm. We have argued elsewhere that a clinical investigator has both an ethical and legal obligation to ensure that individual research subjects will not be exposed to undue risk by study participation.[44] This duty is not fulfilled by merely ensuring that a subject passes each of the eligibility criteria in the study protocol. The clinical researcher must assess carefully the medical history, physical findings, and relevant laboratory results of each prospective subject before making the *clinical determination* that he or she is fit to enter the study. George makes the intriguing argument that large numbers of "safety" criteria may distract clinical researchers from this crucial task:

> A detailed list of safety-type exclusions can paradoxically lead to less attention to other specific details of the individual patient, and this lack of attention can have disastrous consequences. An

> otherwise eligible patient for a clinical trial in cancer who had just been seriously injured in an automobile accident would almost certainly be immediately excluded from further consideration even though this situation was not specifically mentioned as an exclusion criterion. However, another patient with a complicated set of comorbid conditions that leads to an undue risk might be entered confidently, but erroneously, if the (presumed) safety eligibility checklist is met. The key point is that the clinical investigator must, in all cases, make a judgement about the suitability of each patient for entry onto the trial based on all relevant medical and other considerations in addition to the checklist of eligibility requirements. Viewed this way, the eligibility requirements serve as additional and often unnecessary roadblocks to otherwise appropriate patients.[32]

Ultimately, we believe that the NSABP and POG clinical trials represent differing philosophies regarding the design and conduct of clinical cancer research. In their classic paper, Schwarz and Lellouch make a distinction between pragmatic and explanatory clinical research.[45] Explanatory research studies aim at answering biological questions and may, therefore, utilize "strict selection" of eligible research subjects. Pragmatic trials, on the other hand, involve questions relating directly to clinical practice and, thus, a "heterogeneous population" of study

patients is required. We believe that the NSABP trials embrace an explanatory-trial philosophy. As we have seen, the NSABP studies have restricted the eligible patient population for their studies with a variety of criteria that we have classified in precision and safety categories. While such criteria may produce a more homogenous population for study, this comes at the expense of clinical applicability. The POG trials clearly embrace the philosophy of the pragmatic trial. Few precision criteria were added over time and no safety criteria were present in any of the studies. While patient populations in their studies are certainly more heterogeneous, the results of such trials are broadly applicable. Numerous factors undoubtedly affect the proportion of patients treated in clinical trials. Among these factors, differing philosophies of trial design appears to be important: while only 3.3 to 8% of breast cancer patients are treated in trials,[1,46] 79% of children with acute lymphocytic leukemia are treated in research studies.[47]

Recommendations

Based on our findings, we have the following recommendations for the design of clinical trials.

● **The distinction between inclusion and exclusion criteria ought to be abandoned.** Clearly, nothing is gained by distinguishing between inclusion and exclusion criteria: the two categories of criteria are interchangeable by merely

adding a 'not' to any given criterion. As we have seen, when the distinction is made, duplications occur. Such duplications make IRBs and perhaps even clinician-investigators themselves wonder whether the protocol was prepared thoughtfully.

● **Eligibility criteria in clinical trials protocols should be explicitly justified.** The absence of transparent reasoning for each eligibility criterion forced our group to guess as to the intention of trial designers. But the inclusion of a rationale for criteria is of importance to more parties than just those doing research on eligibility criteria. As we have seen, Institutional Review Boards are paying closer attention to criteria. Explanations for individual criteria will help assure IRBs that each criterion has a sound and legitimate basis (as opposed to a frivolous or illegitimate basis).

● **In a series of clinical trials by the same cooperative group, the continuing need for each criterion ought to be assessed when each new trial is planned.** Our "textual analysis" of criteria in the NSABP subset showed that after criteria are added to a series of studies, they usually remain. By scrutinizing carry-over criteria more closely, trial designers can minimize the number of criteria thereby simplifying the patient-enrollment process.

● **Eligibility criteria in phase III clinical trials should restrict the eligible patient population as little as possible, consonant with the demands of**

scientific validity. The contrast between NSABP and POG studies suggests that large numbers of criteria may not be necessary for good quality clinical trials. Minimizing restrictive precision and safety criteria tends to increase the eligible population for the study and makes it more representative. As a result, accrual rates to studies will be enhanced and study findings may have a greater impact on clinical practice.

- **Finally, research is needed to define the precise impact of eligibility criteria on generalizability.** What proportion of patients in a target population is excluded by common eligibility criteria? Do clinician-investigators interpret relatively subjective and objective eligibility criteria differently? How is the impact of study results on clinical practice affected by criteria? Are eligibility criteria reported faithfully in publications of clinical trial results? Do clinicians apply study results to patients who would have been ineligible for study participation? The answers to these questions await further research.

Acknowledgements

The authors are grateful to Dr. Richard Margolese and Ms. Riva Kaplan of the S.M.B.D. Jewish General Hospital for providing copies of the NSABP protocols, and to Ms. Susan Devine of the Montreal Children's Hospital for providing copies of POG protocols. Dr. James Robbins of the Department of Pediatrics, University of Arkansas for Medical Sciences participated in discussions when the study was planned.

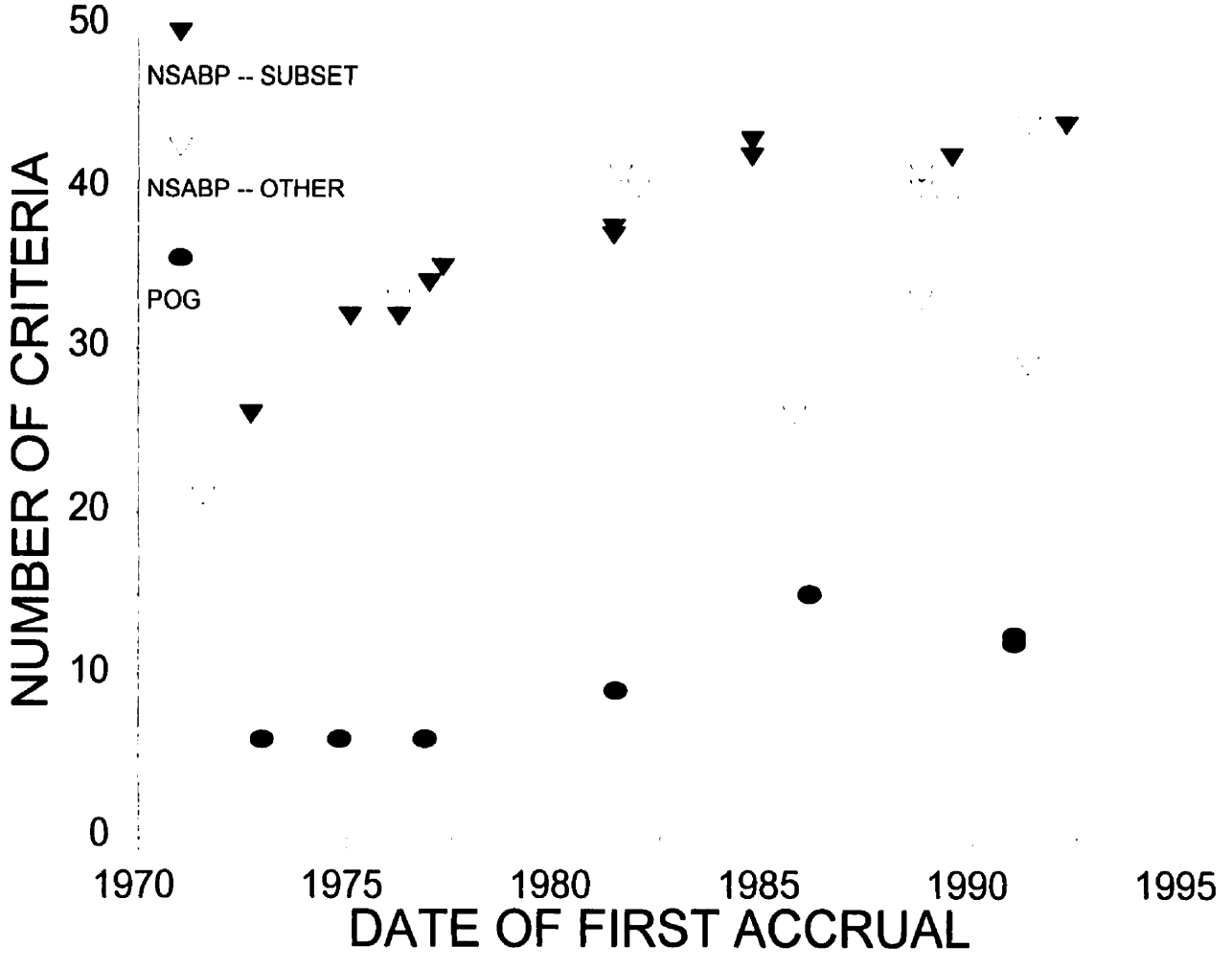

50
40
30
20
10
0
NSABP -- SUBSET
NSABP -- OTHER
POG
NUMBER OF CRITERIA
1970
1975
1980
1985
1990
1995
DATE OF FIRST ACCRUAL

Figure 1. Eligibility criteria in NSABP and POG clinical trials, 1972 to 1992. The total number of eligibility criteria in NSABP (n=22) and POG (n=7) studies (criteria duplications excluded) is plotted against the date of first accrual for each study. NSABP studies are divided into two groups: NSABP-subset studies (n=11) are clinical trials examining chemotherapy in the treatment of node-positive breast cancer; and all other NSABP breast cancer studies (n=11).

Table 1.

Study	Year of first accrual	Total criteria	Classification schema				
			Definition of disease	Precision	Safety	Ethical and legal	Administrative
B-05	1972	26	11	9	3	2	1
B-07	1975	32	13	10	4	4	1
B-08	1976	32	12	11	4	4	1
B-09	1977	34	13	12	4	4	1
B-10	1977	35	13	11	6	4	1
B-11	1981	37	16	13	4	3	1
B-12	1981	37	16	13	4	3	1
B-15	1984	43	18	17	4	3	1
B-16	1984	42	18	16	4	3	1
B-22	1989	42	18	16	4	3	1
B-25	1992	44	18	17	5	3	1
	Mean ± SE	36.72 ± 1.70	15.09 ± 0.83	13.18 ± 0.87	4.18 ± 0.23	3.27 ± 0.19	1.00 ± 0.00
	Correlation (p-value)	0.93 (0.0001)	0.88 (0.0004)	0.88 (0.0003)	0.40 (0.13)	-0.33 (0.21)	--
	Regression coefficient (CI_{95})		0.41 (0.28, 0.54)	0.44 (0.32, 0.56)	0.03 (-0.06, 0.12)	-0.03 (-0.10, 0.05)	--

Table 1. Eligibility criteria in NSABP-subset clinical trials.

Table 2.

Study	Year of first accrual	Total criteria	Classification schema				
			Definition of disease	Precision	Safety	Ethical and legal	Administrative
ALinC-10	1972	6	2	2	0	1	1
ALinC-11	1974	6	2	2	0	1	1
ALinC-12	1976	6	2	2	0	1	1
ALinC-13	1981	9	4	2	0	1	2
ALinC-14	1986	15	5	4	0	1	5
ALinC-15a	1991	12	5	3	0	1	3
ALinC-15b	1991	12	5	3	0	1	3
	Mean ± SE	9.43 ± 1.38	3.57 ± 0.57	2.57 ± 0.30	0.00 ± 0.00	1.00 ± 0.00	2.29 ± 0.57
	Correlation (p-value)	0.71 (.04)	0.87 (0.01)	0.60 (0.08)	--	--	0.71 (0.04)
	Regression coefficient (CI_{95})		0.19 (0.12, 0.26)	0.08 (-0.004, 0.16)	--	--	0.15 (0.005, 0.30)

Table 2. Eligibility criteria in POG clinical trials.

Figure 2.

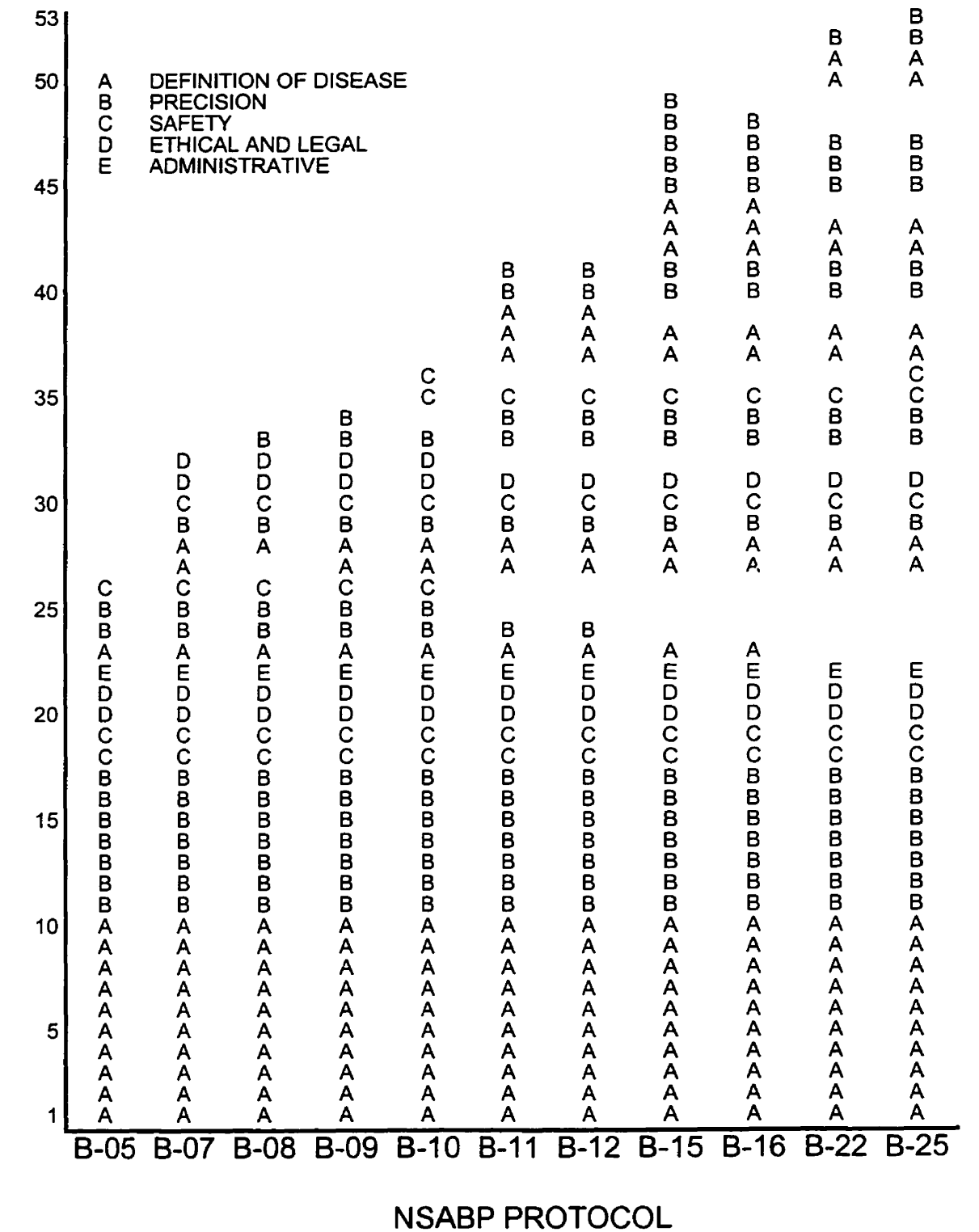

CRITERION
A DEFINITION OF DISEASE
B PRECISION
C SAFETY
D ETHICAL AND LEGAL
E ADMINISTRATIVE
53
50
45
40
35
30
25
20
15
10
5
1
B-05
B-07
B-08
B-09
B-10
B-11
B-12
B-15
B-16
B-22
B-25
NSABP PROTOCOL

Figure 2. Tracking individual eligibility criteria in the series of NSABP-subset clinical trials. The figure tracks the appearance and disappearance of the 53 discrete eligibility criteria found in the eleven NSABP-subset studies. The classification of each criterion is indicated by a letter of the alphabet: A - definition of disease, B - precision, C - safety, D - ethical and legal, E - administrative. The full text of each criterion at entry into the series is given is appendix 1 (the text of a criterion may have been modified slightly over the series).

References

1.Friedman MA, Cain DF. National cancer institute sponsored cooperative clinical trials. Cancer 1990; 65: 2376-2382.

2.American Medical Association Council on Scientific Affairs. Viability of cancer clinical research: patient accrual, coverage, and reimbursement. Journal of the National Cancer Institute 1991; 83: 254-259.

3.Wittes RE, Friedman MA. Editorial: Accrual to clinical trials. Journal of the National Cancer Institute 1988; 80: 884-885.

4.Frelick RW. The Community Clinical Oncology Program (CCOP) story: review of community oncologists' experiences with clinical research trials in cancer with an emphasis on the CCOP of the National Cancer Institute between 1982 and 1987. Journal of Clinical Oncology 1994; 12: 1718-1723.

5.Cobau CD. Clinical trials in the community: the Community Clinical Oncology Program experience. Cancer 1994; 74(supp.): 2694-2700.

6.Cottman CA. The SWOG experience with CCOP's. Journal of Cancer Program Management 1986; 1: 19-22.

7.Begg CB, Carbone PP, Elson PJ, Zelen M. Participation of community hospitals in clinical trials: analysis of five years of experience in the Eastern Cooperative Oncology Group. New England Journal of Medicine 1982; 306: 1076-1080.

8.Gotay CC. Accrual to cancer research trials: directions from the research literature. Social Science and Medicine 1991; 33: 569-577.

9.Taylor KM, Margolese RG, Soskolne CL. Physician's reasons for not entering eligible patients in a randomized clinical trial of surgery for breast cancer. New England Journal of Medicine 1984; 310: 1363-1367.

10.Mansour EG. Barriers to clinical trials. Part III: Knowledge and attitudes of health care providers. Cancer 1994; 74(supp.): 2672-2675.

11.Benson AB III, Pregler JP, Bean JA, Rademaker AW, Eshler B, Anderson K. Oncologists' reluctance to accrue patients onto clinical trials: an Illinois Cancer Center study. Journal of Clinical Oncology 1991; 9: 2067-2075.

12.Taylor KM, Feldstein ML, Skeel RT, Pandya KJ, Ng P, Carbone PP. Fundamental dilemmas of the randomized clinical trial process: results of a survey of the 1,737 Eastern Cooperative Oncology Group investigators. Journal of Clinical Oncology 1994; 12: 1796-1805.

13.Langley GR, Sutherland HJ, Wong S, Minkin S, Llewellyn-Thomas HA, Till JE. Why are (or are not) patients given the option to enter clinical trials? Controlled Clinical Trials 1987; 8: 49-59.

14.Morrow GR, Hickok JT, Burish TG. Behavioral aspects of clinical trials: an integrated framework from behavior theory. Cancer 1994; 74(supp.): 2676-2682.

15.Llewellyn-Thomas HA, McGreal MJ, Thiel EC, Fine S, Erlichman C. Patients' willingness to enter clinical trials: measuring the association between perceived benefit and preference for decision participation. Social Science and Medicine 1991; 32: 35-42.

16.Schain WS. Barriers to clinical trials. Part II: Knowledge and attitudes of potential participants. Cancer 1994; 74 (supp.): 2666-2671.

17.Llewellyn-Thomas HA, McGreal MJ, Theil EC. Cancer patients' decision making and trial-entry preferences. Medical Decision Making 1995; 15: 4-12.

18.Cassileth BR, Lusk EJ, Miller DS, and colleagues.. Attitudes towards clinical trials among patients and the public. JAMA 1982; 248: 968-970.

19.McCusker J, Wax A, Bennett JM. Cancer patient accessions into clinical trials: a pilot investigation into some patient and physician determinants of entry. American Journal of Clinical Oncology 1982; 5: 227-236.

20.Kotwall CA, Mahoney LJ, Myers RE, Decoste L. Reasons for nonentry in randomized clinical trials for breast cancer: a single institution study. Journal of Surgical Oncology 1992; 50: 125-129.

21.Begg CB, Zelen M, Carbone PP, and colleagues.. Cooperative groups and community hospitals: measurement of impact in the community hospitals. Cancer 1983; 52: 1760-1767.

22.Lee JY, Breaux SR. Accrual of radiotherapy patients to clinical trials. Cancer 1983; 52: 1014-1016.

23.Weijer C. The breast cancer research scandal: addressing the issues. Canadian Medical Association Journal 1995; 152: 1195-1197.

24.Boros L, Chuange C, Butler FO, Bennett JM, Leukemia in Rochester: a 17-year-experience with an analysis of the role of cooperative group (ECOG)

participation. Cancer 1985; 56: 2161-2169.

25.Karjalainen S, Palva I. Do treatment protocols improve end results? A study of the survival of patients with multiple myeloma in Finland. British Medical Journal 1989; 299: 1069-1072.

26.Hjorth M, Holmberg E, Rodjer S, Westin J. Impact of active and passive exclusions on the results of a clinical trial in multiple myeloma. British Journal of Haematology 1992; 80: 55-61.

27.Weijer C, Freedman B, Fuks A, Robbins J, Shapiro S, Skrutkowska M. What difference does it make to be treated on a clinical trial? -- A pilot study. Clinical and Investigative Medicine 1996; 19: 179-183.

28.Yusuf S, Collins R, Peto R. Why do we need large, simple trials? Statistics in Medicine 1984; 3: 409-420.

29.Sackett DL. On some prerequisites for a successful clinical trial. In: Shapiro S, Louis TA (eds.). Clinical Trials: Issues and Approaches. New York: Marcel Dekker, Inc., 1983: 65-79.

30.Yusuf S, Held P, Teo KK. Selection of patients for randomized controlled trials: implications of wide or narrow eligibility criteria. Statistics in Medicine 1990; 9: 73-86.

31.Buyse ME. The case for loose exclusion criteria in clinical trials. Acta Chir Belg 1990; 90: 129-131.

32.George SL. Reducing patient eligibility criteria in cancer clinical trials. Journal

of Clinical Oncology 1996; 14: 1364-1370.

33.Begg CB, Engstrom PF. Eligibility and extrapolation in cancer clinical trials. Journal of Clinical Oncology 1987; 5: 962-968.

34.Fleiss JL. Statistical Methods for Rates and Proportions. 2nd ed.. New York: John Wiley & Sons, 1981: 212-236.

35.The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. OPRR Reports; April 18, 1979: 1-8.

36.Weijer C. Evolving ethical issues in the selection of subjects for clinical research. Cambridge Quarterly of Healthcare Ethics 1996; 5: 334-345.

37.Department of Health and Human Services, National Institutes of Health. NIH Guidelines on the inclusion of women and minorities as subjects in clinical research. Federal Register 1994; 59: 14508-14513.

38.Freedman B. Mutlicenter trials and subject eligibility: should local IRBs play a role? IRB: A Review of Human Subjects Research 1994; 16(1,2): 1-6.

39.Kennedy BJ. Needed: clinical trials for older patients. Journal of Clinical Oncology 1991; 9: 718-720.

40.Mor V, Masterson-Allen S, Goldberg RJ, Cummings FJ, Glicksman AS, Fretwell MD. Relationship between age at diagnosis and treatments received by cancer patients. Journal of the American Geriatric Society 1985; 33: 585-589.

41.Samet J, Hunt WC, Key C, Humble CG, Goodwin JS. Choice of cancer therapy varies with age of patient. JAMA 1986; 255: 3385-3390.

42.Giovanazzi-Bannon S, Rademaker A, Lai G, Benson AN. Treatment tolerance of elderly cancer patients entered onto phase II clinical trials: an Illinois Cancer Center study. Journal of Clinical Oncology 1994; 12: 2447-2452.

43.Christman K, Muss HB, Case LD, Stanley V. Chemotherapy of metastatic breast cancer in the elderly: the Piedmont Oncology Association experience. JAMA 1992; 268: 57-62.

44.Weijer C, Fuks A. The duty to exclude: excluding people at undue risk from research. Clinical and Investigative Medicine 1994; 17: 115-122.

45.Schwarz D, Lellouch J. Explanatory and pragmatic attitudes in therapetic trials. Journal of Chronic Disease 1967; 20: 637-648.

46.Tate HC, Rawlinson JR. Randomized comparative studies in the treatment of cancer in the United Kingdom: room for improvement? Lancet 1979; ii: 623-625.

47.Meadows AT, Kramer S, Hopson R, Lustbader E, Jarret P, Evans AE. Survival in childhood acute lymphocytic leukemia: effect of protocol and place of treatment. Cancer Investigation 1983; 1: 49-55.

Appendix 1. Full text of eligibility criteria in NSABP-subset clinical trials.

The full text of each of the 53 eligibility criteria in the NSABP studies is listed here. The text is taken from the protocol in which the criterion first appeared. Minor alterations in the text of the criterion may have occurred in subsequent protocols. The numbering for each criterion corresponds to figure 2.

Criteria present in all NSABP-subset studies

1. There is no edema of the arm.
2. The axillary nodes are movable in relation to the chest wall and neurovascular bundle.
3. The tumor is confined to the breast or breast and axilla.
4. The tumour is movable in relation to the underlying muscle and the chest wall.
5. Patients with skeletal pain are considered curable if roentgen examination or scan fails to reveal metastatic disease. Scans are not required; however, if they demonstrate metastases in patients having a positive finding at the site of pain, that patient is not eligible for the study; but, asymptomatic patients having a positive bone scan will not be considered ineligible for this protocol unless there is biopsy or roentgen proof of metastases.

6. Patients with malignant breast tumors other than carcinoma [are excluded].

7. Patients with findings which relegate them to the category of inoperability such as peau d'orange involving greater than 1/3 of the breast, satellite or parasternal nodules [are excluded].

8. Patients with significant nodes in the opposite axilla or palpable supraclavicular nodes are considered incurable unless there is biopsy proof that these nodes are uninvolved [are excluded].

9. Patients with bilateral malignancy [are excluded].

10. Having one (1) or more positive axillary nodes proven histologically.

11. Patients with previous oophorectomy (surgical or radiation castration) are eligible for study if the oophorectomy was not performed for tumor.

12. Patients with a previous malignancy, regardless of the site [are excluded] - EXCEPT patients with squamous or basal cell carcinoma of the skin which can be effectively treated.

13. Patients with a concomitant malignancy, regardless of site [are excluded] - EXCEPT patients with squamous or basal cell carcinoma of the skin which can be effectively treated.

14. Patients who are poor surgical risks having non-malignant systemic disease (cardiovascular, renal, etc.) which would preclude their being subject to any of the treatment options and who are at high-risk for prolonged follow-up

[are excluded].

15. Prior irradiation, surgery, or chemotherapy for breast cancer [necessitates exclusion].

16. Prior hormonal therapy for breast cancer [necessitates exclusion].

17. Had a radical mastectomy (conventional or modified).

18. WBC $\geq$ 4,00/cu.mm. and Platelet count $\geq$ 100,00/cu.mm.

19. BUN $\leq$ 25 mg%.

20. Patient consents to be included in the study.

21. Patients who are pregnant [are excluded].

22. Patient is accessible (geographically) for follow-up.

Criteria present in B-05, but not all subsequent protocols

23. Female patients.

24. Patients over 75 years of age [are excluded].

25. Patients who have been treated previously for their current malignancy including those whose previous "treatment" has been biopsy only (including excisional biopsy) which was performed more than four weeks prior to radical mastectomy [are excluded].

26. Patients who are lactating (includes those patients who have stopped lactating in the past 6 months) [are excluded].

Criteria added in B-07 and subsequent protocols

27. Ipsilateral axillary lymph nodes over 2 cm in greatest diameter [mandate exclusion].

28. Inflammatory carcinoma [necessitates exclusion].

29. Therapy must begin 2-4 weeks after mastectomy.

30. Evidence of adequate hepatic function (bilirubin $\leq$ 1.5 mg%, SGOT $\leq$ 60 I.U./ml).

31. Patients with psychiatric or addictive disorders which would preclude obtaining informed consent [are excluded].

32. Patients with psychiatric or addictive disorders which would preclude their being subject to any of the treatment options [are excluded].

33. Prior therapy for breast cancer, including prior immunotherapy for breast cancer [necessitates exclusion].

34. Tumor of patient is available for estrogen/progesterone receptor site analyses. (Patients whose tumors are too small for receptor analysis as documented by pathology report will be accepted and followed.)

35. Patients with a history of hypertension, coronary artery disease, previous myocardial infarction or any other cardiovascular disease [are excluded].

36. Patients with hypersensitivity to any protein material or pre-existing auto-immune disorder [are excluded].

37. The overlying skin must be movable with respect to the tumor.

38. The tumor is invasive on histological examination.

39. Having one or more tumors negative for estrogen receptors. Estrogen receptor data must be reported quantitatively in fmol/mg cytosol protein and must be < 10 fmol/mg cytosol.

40. Patients receiving any hormonal therapy other than that stipulated in the protocol, eg. birth control pills, replacement therapy, etc., are eligible if this therapy is discontinued while on protocol therapy.

41. Patients whose histologic diagnosis has been established more than four weeks prior to mastectomy [are excluded]. This includes excisional, incisional or needle biopsy and aspiration cytology.

42. On clinical examination, the tumor was 4 cm. or less in its greatest dimension (mammographic measurements should be used when possible).

43. Patients treated with segmental mastectomy in whom there is another dominant mass within the ipsilateral breast remnant [are excluded]. Such a mass must be biopsied and demonstrated to be histologically benign prior to randomization.

44. Patients who have met the age and receptor criteria:

AGE	RECEPTOR STATUS
< 49	All patients regardless of receptor status
50-59	Patients with tumor PR < 10fmol regardless of ER level.

45. Patients who received breast radiation therapy following segmental mastectomy prior to randomization[are excluded].

46. Patients must have a life expectancy of at least 10 years excluding their diagnosis of cancer.

47. The breast was of sufficient size to permit a cosmetically acceptable resection.

48. Patients with breasts deemed too large to permit satisfactory radiation to be delivered[are excluded].

49. Patients treated with segmental mastectomy in whom the nipple was removed [are excluded].

50. Patients with any distant metastasis [are excluded].

51. Patients with diffuse tumors as demonstrated on xeroradiography or mammography which would not be considered surgically amenable to lumpectomy [are excluded].

52. The margins of the resected specimen must be histologically free of invasive and non-invasive tumor. In patients where pathologic examination demonstrates tumor present at the line of resection, one additional operative procedure may be performed to obtain clear margins. This is permissible even if axillary dissection has been performed. Patients in whom tumor is still present after the second resection must undergo total mastectomy.

53. Patients who have undergone a radical mastectomy (removal of the breast and complete removal of the pectoralis major muscle [are excluded]. Partial excision of the muscle does not constitute a radical mastectomy.

Appendix 2: Inter-rater reliability questionnaire.

Letter to questionnaire recipients

Dear Colleague:

The Clinical Trials Research Group, as a part of its research on eligibility criteria and clinical trials, has developed a schema for classifying eligibility criteria used in clinical trials. We are requesting your assistance in assessing the inter-rater reliability of our schema.

Our schema classifies eligibility criteria into five broad categories according to their purpose (a sixth catch-all "other" category is added for completeness). Criteria are classified into the following categories: (A) definition of disease; (B) precision; (C) safety; (D) legal and ethical; and, (E) administrative. Each of these categories is defined on the following page. If a given criterion is equally well described by two categories, the category closer to the beginning of the list is chosen (e.g., if a criterion is both "precision" and "safety", then "precision" is chosen).

We would like you to classify the criteria from a recent NSABP adjuvant breast cancer protocol. We have provided a brief overview of the study for you to place the criteria in a context. For each criterion, we would like you to select the one category that best describes it. (We have not included the "other" category as an option as we wish to "force" a choice).

We estimate this "exercise" will occupy twenty minutes of your time and we are grateful for your help. If you do not wish to participate, please do not fill out the questionnaire. Your answers to the questionnaire will, of course, be kept confidential.

Thank you for your kind assistance!

The Clinical Trials Research Group

Benjamin Freedman, Ph.D.

Abe Fuks, M.D., C.M., F.R.C.P.(C)

Stan Shapiro, Ph.D.

Kathy Glass, D.C.L.

Charles Weijer, M.D.

Myriam Skrutkowska, B.Sc.N. (Research coordinator)

ELIGIBILITY CRITERIA CLASSIFICATION SCHEMA:

(A) DEFINITION OF DISEASE

The purpose of definitional criteria is to set a defined study population of disease. These criteria mirror the factors that would be taken into account in the clinical setting, e.g., stage of disease, pathologic sub-type, etc.. E.g.,

[Exclude] Patients with tumors greater than 5 cm in size at the largest dimension on clinical examination.

(B) PRECISION (POPULATION AND MEASUREMENT)

These criteria, unique to research protocols, are concerned with the scientific validity of the research study. These criteria diminish variability in the study by either making the patient population more homogenous or by reducing measurement error. Typically, these are factors that would not influence treatment decisions in clinical practice, e.g., prior history of cancer, prior treatment with study agents, etc.. E.g.,

[Exclude] Patients who have had prior therapy for their breast cancer, including irradiation and chemotherapy.

(C) SAFETY

Criteria in this category serve the purpose of protecting the vulnerable subject from the risks of treatment in general or from the risks of specific treatments in the particular research protocol. These criteria serve to identify and exclude patients who are at a higher risk of developing ill-effects from the study treatments. E.g.,

There must be evidence postoperatively of adequate hepatic functions (bilirubin within normal limits, SGOT or SGPT within normal limits).

(D) LEGAL AND ETHICAL

Criteria in this category ensure the compliance of the protocol with the regulatory (MRC Guidelines; DHHS Regulations) and/or legal requirements for human experimentation. E.g.,

Patients must consent to be in the study. The informed consent must be signed, witnessed, and dated prior to randomization.

(E) ADMINISTRATIVE

Administrative criteria set out to ensure the smooth functioning of the mechanics of the clinical trial. Criteria aimed at ensuring compliance with treatment and follow-up fall into this category. E.g.,

Patients must be geographically accessible for follow-up.

NSABP PROTOCOL NO. B-22

A CLINICAL TRIAL TO EVALUATE THE EFFECT OF DOSE INTENSIFICATION AND INCREASED CUMULATIVE DOSE OF POST-OPERATIVE ADRIAMYCIN- CYCLOPHOSPHAMIDE (AC) THERAPY ON THE DISEASE-FREE SURVIVAL AND SURVIVAL OF PATIENTS WITH PRIMARY BREAST CANCER AND POSITIVE AXILLARY NODES.

Overview

Patients who have histologically proven primary operable breast cancer with one or more histologically positive axillary lymph nodes and no evidence of metastatic disease will be eligible for this study. Patients will be stratified by number of positive nodes, age, ER level and type of operation and then randomly assigned to one of three treatment groups (see figure). Patients in all groups will undergo either (a) lumpectomy plus axillary dissection followed by breast radiation after completion of all chemotherapy or (b) total mastectomy. Following operation, patients in group I will be treated with AC therapy, i.e., four cycles (courses) of Adriamycin (60 mg/m^2) and cyclophosphamide (CY; 600 mg/m^2), with a 21 day interval between courses. Patients in group II will receive the same

dose of Adriamycin (60 mg/m^2) at 21-day intervals X 4, just as in group I. CY, however, will be 1200 mg/m^2, administered for two courses, i.e., on days 1 and 22, with the Adriamycin. No CY will be given during courses 3 and 4. Patients in group II will thus receive the same cumulative dose of CY as those in group I, but all CY will be given in two "intensified" courses. Patients in group III will receive an intensified dose of CY plus a greater cumulative (total) dose to be administered AC [Adriamycin (60 mg/m^2) and cyclophosphamide (1200 mg/m^2)] q21 days X 4.

In all three groups, patients who are ≥50 years will receive tamoxifen, 10 mg bid, beginning on Day 1 of Cycle 1, and will continue this regimen for five years.

A - DEFINITION OF DISEASE
B - PRECISION (OF POPULATION AND MEASUREMENT)
C - SAFETY
D - LEGAL AND ETHICAL
E - ADMINISTRATIVE

1.(Exclude) Patients treated with lumpectomy in whom there is another dominant mass within the ipsilateral breast remnant. Such a mass must be biopsied and demonstrated to be histologically benign prior to randomization.

A B C D E

2.(Exclude) Patients who receive breast radiation therapy following lumpectomy but prior to randomization.

A B C D E

3.Patients must have a life expectancy of at least 10 years, excluding their diagnosis of cancer.

A B C D E

4.(Exclude) Patients who have had prior therapy for their breast cancer, including hormonal therapy.

A B C D E

5.(Include) (Patients) Treated by lumpectomy and axillary dissection or total mastectomy and axillary dissection.

A B C D E

Please make only one choice for each criterion. If two categories seem equally good descriptions, choose the one closest to the beginning of the alphabet. Please make a choice for every criterion even if you are not sure of a given choice.

A - DEFINITION OF DISEASE
B - PRECISION (OF POPULATION AND MEASUREMENT)
C - SAFETY
D - LEGAL AND ETHICAL
E - ADMINISTRATIVE

6.The postoperative WBC must be ≥4 000/cu mm and platelet count ≥100 000/cu mm.

A B C D E

7.There must be evidence postoperatively of adequate renal function (serum creatinine ≤ 1.5 mg%).

A B C D E

8.(Exclude) Patients with bilateral malignancy or a mass in the opposite breast unless there is biopsy proof that the mass is not malignant.

A B C D E

9.On clinical examination, axillary nodes, if palpable, must be movable in relation to the chest wall and neurovascular bundle.

A B C D E

10.The tumor must be confined to the breast or breast and ipsilateral axilla.

A B C D E

Please make only one choice for each criterion. If two categories seem equally good descriptions, choose the one closest to the beginning of the alphabet. Please make a choice for every criterion even if you are not sure of a given choice.

A - DEFINITION OF DISEASE
B - PRECISION (OF POPULATION AND MEASUREMENT)
C - SAFETY
D - LEGAL AND ETHICAL
E - ADMINISTRATIVE

11.The overlying skin must be moveable with respect to the tumor.

A B C D E

12.The tumor must be invasive on histologic examination.

A B C D E

13.Patients receiving any sex hormonal therapy other than that stipulated in the protocol, e.g., birth-control pills, ovarian hormonal replacement therapy, etc., are eligible if this therapy is discontinued while on protocol (until first treatment failure).

A B C D E

14.(Exclude) Patients with ipsilateral axillary lymph nodes which are greater than 2 cm in greatest diameter.

A B C D E

15.(Exclude) Patients with inflammatory carcinoma.

A B C D E

Please make only one choice for each criterion. If two categories seem equally good descriptions, choose the one closest to the beginning of the alphabet. Please make a choice for every criterion even if you are not sure of a given choice.

A - DEFINITION OF DISEASE
B - PRECISION (OF POPULATION AND MEASUREMENT)
C - SAFETY
D - LEGAL AND ETHICAL
E - ADMINISTRATIVE

16.(Exclude) Patients whose histologic diagnosis has been established more than 28 days prior to mastectomy. Diagnosis includes excisional, incisional or needle biopsy and aspiration cytology.

A B C D E

17.(Exclude) Patients with active cardiac disease that would preclude the use of Adriamycin. This includes:

(a) Any documented myocardial infarction.

(b) Angina pectoris which requires the use of anti-anginal medication.

(c) Any history of documented congestive heart failure.

(d) Patients with a cardiac arrhythmia are eligible for this protocol, provided the arrhythmia is not associated with concomitant heart failure or cardiac dysfunction.

(e) Valvular disease with documented cardiac function compromise.

(f) Cardiomegaly on chest x-ray.

(g) Poorly controlled hypertension, i.e., diastolic greater than 100mm/Hg. Those patients with hypertension that is well controlled on medication are eligible for entry.

A B C D E

Please make only one choice for each criterion. If two categories seem equally good descriptions, choose the one closest to the beginning of the alphabet. Please make a choice for every criterion even if you are not sure of a given choice.

A - DEFINITION OF DISEASE
B - PRECISION (OF POPULATION AND MEASUREMENT)
C - SAFETY
D - LEGAL AND ETHICAL
E - ADMINISTRATIVE

18.On clinical examination, the tumor must be 5 cm or less in its greatest dimension.

A B C D E

19.Patients must consent to be in the study. The informed consent form must be signed, witnessed, and dated prior to randomization.

A B C D E

20.Patients with palpable nodes in the axilla opposite the affected breast or with palpable supraclavicular or infraclavicular nodes are considered ineligible unless there is biopsy evidence that these are not involved with tumor.

A B C D E

21.The tumor must be movable in relation to the underlying muscle and chest wall on clinical examination, i.e., not attached to pectoral fascia or chest wall.

A B C D E

22.The breast must be of sufficient size to permit a cosmetically acceptable resection.

A B C D E

23.(Exclude) Patients with any distant metastasis.

A B C D E

Please make only one choice for each criterion. If two categories seem equally good descriptions, choose the one closest to the beginning of the alphabet. Please make a choice for every criterion even if you are not sure of a given choice.

A - DEFINITION OF DISEASE
B - PRECISION (OF POPULATION AND MEASUREMENT)
C - SAFETY
D - LEGAL AND ETHICAL
E - ADMINISTRATIVE

24.(Exclude) Patients with diffuse tumors as demonstrated on xeroradiography or mammography which would not be considered surgically amenable to lumpectomy.

A B C D E

25.(Exclude) Patients who are pregnant at the time of randomization.

A B C D E

26.The interval between the definitive operation and date of first treatment must be no more than 35 days.

A B C D E

27.There must be evidence postoperatively of adequate hepatic functions (bilirubin ≤1.5%, SGOT or SGPT ≤60 IU/ml).

A B C D E

28.(Exclude) Patients with psychiatric or addictive disorders that would preclude obtaining informed consent.

A B C D E

Please make only one choice for each criterion. If two categories seem equally good descriptions, choose the one closest to the beginning of the alphabet. Please make a choice for every criterion even if you are not sure of a given choice.

A - DEFINITION OF DISEASE
B - PRECISION (OF POPULATION AND MEASUREMENT)
C - SAFETY
D - LEGAL AND ETHICAL
E - ADMINISTRATIVE

29.(Exclude) Patients who have had prior therapy for their breast cancer, including immunotherapy.

A B C D E

30.There must be no edema of the arm.

A B C D E

31.(Exclude) Patients with a concomitant malignancy, regardless of site, except patients with squamous or basal cell carcinoma of the skin which has been effectively treated or carcinoma in situ of the cervix or uterus which has been treated by operation only.

A B C D E

32.(Exclude) Patients who have nonmalignant systemic disease (cardiovascular, renal, hepatic, etc.), which would preclude their being subjected to any of the treatment options or would prevent prolonged follow-up.

A B C D E

Please make only one choice for each criterion. If two categories seem equally good descriptions, choose the one closest to the beginning of the alphabet. Please make a choice for every criterion even if you are not sure of a given choice.

A - DEFINITION OF DISEASE
B - PRECISION (OF POPULATION AND MEASUREMENT)
C - SAFETY
D - LEGAL AND ETHICAL
E - ADMINISTRATIVE

33.(Exclude) Patients who have had prior therapy for their breast cancer, including irradiation or chemotherapy.

A B C D E

34.Patients with skeletal pain are eligible for inclusion in the study if bone scan and/or roentgenological examination fail to disclose metastatic disease.

A B C D E

35.(Exclude) Patients with breast tumors other than carcinoma.

A B C D E

36.Patients must be accessible geographically for follow-up.

A B C D E

37.At least one axillary lymph node must demonstrate evidence of tumor on histologic examination.

A B C D E

Please make only one choice for each criterion. If two categories seem equally good descriptions, choose the one closest to the beginning of the alphabet. Please make a choice for every criterion even if you are not sure of a given choice.

A - DEFINITION OF DISEASE
B - PRECISION (OF POPULATION AND MEASUREMENT)
C - SAFETY
D - LEGAL AND ETHICAL
E - ADMINISTRATIVE

38.Patients with previous surgical oophorectomy are eligible for this study if the oophorectomy was not performed for malignancy. (Radiation castration will render the patient ineligible).

A B C D E

39.(Exclude) Patients with a previous malignancy, regardless of site, except patients with squamous or basal cell carcinoma of the skin which has been effectively treated or carcinoma in situ of the cervix or uterus which has been treated by operation only.

A B C D E

40.(Exclude) Patients with findings which relegate them to the category of more advanced disease, such as peau d'orange or skin edema of any magnitude, satellite breast nodules or parasternal nodules, and edema of the arm. (The term satellite nodules refers to discrete foci of tumor involving the skin around the tumor. Should these be suspected, they must be examined histologically, and, if confirmed, the patient should be considered ineligible for this study).

A B C D E

Please make only one choice for each criterion. If two categories seem equally good descriptions, choose the one closest to the beginning of the alphabet. Please make a choice for every criterion even if you are not sure of a given choice.

A - DEFINITION OF DISEASE
B - PRECISION (OF POPULATION AND MEASUREMENT)
C - SAFETY
D - LEGAL AND ETHICAL
E - ADMINISTRATIVE

41.A quantitative estrogen and progesterone receptor analysis must have been performed in a laboratory which has complied with NSABP quality-control prerequisites. Quantitative estrogen and progesterone receptor information must be available prior to randomization. Estrogen and progesterone receptor data must be reported quantitatively in fmol/mg cytosol protein. No other methodologies for receptor determination will be accepted.

A B C D E

42.The margins of the resected specimen must be histologically free of invasive and non-invasive tumor. In patients where pathologic examination demonstrates tumor at the line of resection, one additional operative procedure may be performed to obtain clear margins. This is permissible even if axillary dissection has been performed. Patients in whom tumor is still present after the second resection must undergo total mastectomy.

A B C D E

Please make only one choice for each criterion. If two categories seem equally good descriptions, choose the one closest to the beginning of the alphabet. Please make a choice for every criterion even if you are not sure of a given choice.

CLASSIFICATION KEY:

1. A
2. B
3. B
4. B
5. B
6. C
7. C
8. A
9. A
10. A
11. A
12. A
13. B
14. A
15. A
16. B
17. C
18. A
19. D
20. A
21. A
22. B
23. A
24. A
25. D
26. B
27. C
28. D
29. B
30. A
31. B
32. B
33. B
34. A
35. A
36. E
37. A
38. B
39. B
40. A
41. B
42. B

Bridging section

In the second chapter, my colleagues and I began the investigation into selection procedures for research participation with an empirical examination of two sets of protocols. Following on the work of Begg and Engstrom, we sought to answer three questions: How have numbers of eligibility criteria in comparable trials changed over time? What is the nature of criteria that have been added or dropped? How do these changes affect the generalizability of study results?

Despite the fact that both the NSABP and POG trials involved anti-cancer chemotherapy, the NSABP trials contained many more eligibility criteria than the POG trials. The number of eligibility criteria approximately doubled in both sets of clinical trials over the twenty-year study interval.

In order evaluate the nature of the added criteria, we developed (and validated) a schema to classify eligibility criteria into five categories: definition of disease, precision, safety, ethical and legal, and administrative. The classification schema was designed with the applicability of study results (generalizability) in mind. Definition of disease criteria are of little concern in this regard; by definition, they represent factors taken into account in clinical practice. The addition of precision criteria, as factors not taken into account in clinical practice, may substantially hamper the applicability of study results. Safety criteria do not hamper generalizability insofar as they mimic prudent clinical decision making.

We found that the increase in eligibility criteria in the NSABP studies was largely accounted for by additions to the definition of disease and precision categories. Furthermore, the NSABP studies added precision criteria at a greater rate than the POG studies. Finally, we were surprised to find that -- as opposed to the NSABP studies -- the POG studies contained no safety criteria whatsoever.

The NSABP and POG clinical trials resulted in important contributions to knowledge in the treatment of breast cancer and acute lymphocytic leukemia. By excluding from study participation groups of patients who would need similar treatment in clinical practice, including older patients, those with prior malignancies, and those with minimal to moderate organ dysfunction, however, the NSABP trials fail to provide information on how best to treat these important groups of patients. The absence of similar criteria in the POG trials suggests that such criteria may not be necessary. This contrast led us to recommend, *inter alia*, that *eligibility criteria in phase III clinical trials should restrict the eligible patient population as little as possible, consonant with the demands of scientific validity.*

At the end of the chapter, we identified a number of areas for further study. In the next chapter, we turn from the protocol to the investigator and address one of these areas directly. We ask: Are subjective eligibility criteria associated with more variable enrollment decisions? Are investigators less certain of their decisions when using such criteria?

Chapter 3:

Measuring the interpretation of criteria for clinical trial eligibility: a survey of 365 oncology investigators.

Abstract

Randomized controlled trials (RCTs) play an important role in ensuring that new medical treatments are both safe and effective. A number of eligibility criteria commonly used in RCTs have been criticized in the literature, including criteria that exclude the elderly, persons with psychiatric disease, or persons with substance abuse problems from trial participation. In this paper, we invoke a novel critique against such criteria: they are subjective, i.e., open to a wide-range of interpretation by RCT investigators. Subjective criteria are, we hypothesise, a source of variability in enrollment decisions and investigator uncertainty. In order to test our hypotheses, we surveyed 365 cancer investigators from the United States and Canada. Investigators were presented with clinical vignettes from three patient categories -- eligible, uncertain, and ineligible -- for each of five eligibility criteria (three subjective and two objective) and asked whether they would enroll the patient in a trial and how sure they are of this decision. Overall, 224 usable questionnaires were returned (response rate = 61.4%). Subjective criteria were associated with more variable enrollment decisions than objective criteria for each of the patient scenarios (eligible scenario, p=0.07; uncertain and ineligible scenarios, p=0.0001). Clinical investigators were also more unsure of decisions made for subjective criteria than objective criteria (all patient scenarios, p=0.001). Demographic characteristics of investigators failed to explain the observed

differences in enrollment decisions or certainty. Subjective eligibility criteria may interfere with the conduct and interpretation of RCTs and their use ought, therefore, to be justified explicitly in the study protocol. Trial designers, funding agencies and Institutional Review Boards have an important role in reviewing eligibility criteria for their necessity. (266 words)

Introduction

In oncology, as in other areas of medicine, randomized clinical trials (RCT) provide us with the most reliable information on the safety and efficacy of novel medical interventions. RCTs provide an unbiased comparison of a new treatment and control treatment (often standard therapy) by assigning patients by chance to the study's differing treatment groups. But to whom do the results of such carefully conducted studies apply? Few patients are actually treated in RCTs, (Tate, 1979; Friedman, 1990) and those that are included represent a highly-select group. In oncology (among the few areas in which good published studies are available), the majority of patients for whom a RCT is available are ineligible for study participation because they do not meet at least one of the (typically) lengthy list of trial eligibility criteria. (Gotay, 1991; McCusker, 1982; Kotwall, 1992; Begg, Zelen and Carbone, 1983; Lee, 1983)

Some specific eligibility criteria have recently come under attack in the literature. Notably, criteria that restrict study populations to men only or exclude members of racial minority groups have been criticized on a number of grounds: excluded individuals are denied the potential benefits associated with RCT participation, the results of such studies may not be applicable to excluded groups, and the resulting gaps in medical knowledge may lead to either inappropriate treatment or under-treatment for women or racial minorities. (Dresser, 1992;

Mastroianni, 1994) In response to these criticisms, the U.S. National Institutes of Health (NIH) issued guidelines in 1994 requiring, in part, the representative inclusion of women and racial minorities in all NIH-funded research.(NIH, 1994)

Similar concerns have been raised regarding other eligibility criteria, including criteria that exclude the elderly from clinical studies. Cancer RCTs funded by the U.S. National Cancer Institute have excluded the elderly from RCTs for years. (Begg and Carbone, 1983) Trials have excluded the elderly directly by setting an age cut-off for trial eligibility (e.g., 70 years of age) and indirectly by excluding persons with co-morbid diseases that are more common in the elderly. Recent breast cancer RCTs of the National Surgical Adjuvant Breast and Bowel Project (NSABP) have abandoned an age cut off in favor of a criterion requiring that "patients must have a life expectancy of at least 10 years excluding their diagnosis of cancer." But how do clinical investigators interpret such a criterion? There is no evidence that more older persons are being enrolled in NSABP RCTs (indeed their representation has declined since the introduction of the new criterion).(Fuks, unpublished data) The exclusion of older persons from cancer research studies has lead to a lack of knowledge about how best to treat cancer in the elderly. (Kennedy, 1991) At least two studies have attributed the systematic under-treatment of cancer in older persons to the exclusion of the elderly from carefully conducted randomized controlled trials. (Mor, 1985; Samet, 1986)

Eligibility criteria excluding persons with a history of psychiatric illness or drug or alcohol abuse have been criticized on other grounds. RCTs often exclude persons with a history of psychiatric disease on the basis that such persons are *at risk* of being or becoming incompetent. Similarly, persons with a history of drug or alcohol abuse are often ineligible for study participation based upon the presumption that they are less compliant than other research subjects. Putting aside the obvious discriminatory tone of such criteria, the associations between psychiatric disease and (*ipso facto*) incompetence and between a history of drug and alcohol abuse and non-compliance can be challenged. Indeed, a patient may have one of any number of psychiatric diagnoses, including major depression, bipolar disorder, and schizophrenia, and remain legally competent. (Michels, 1994; Grisso, 1991) If competency is the issue, then it should be assessed directly. The association between a history of drug and alcohol abuse and non-compliance is even more dubious. Indeed, finding no support for such an association in the empirical literature, Hughes recommends that such eligibility criteria not be used in RCTs. (Hughes, 1993) Again, if compliance is the issue, it is better assessed directly.

We believe that another problem may exist with many of the eligibility criteria mentioned above: they are open to a wide range of interpretation by RCT investigators. Eligibility criteria that require investigators to assess prospective

research subjects on the basis of life expectancy, unacceptable co-morbid diseases, or psychiatric disease associated with incompetency are all examples of subjective criteria that are likely to be interpreted differently by different investigators. Put another way, assessed by the same subjective criterion, a given patient may be deemed eligible in one study centre and ineligible in another. If decisions involving individual criteria are highly variable, such criteria may represent an avoidable source of added variability to the study and, hence, the study's internal validity may be diminished by their presence. Also, highly-subjective criteria may make it more difficult for clinicians reading the published study to decide if the results apply to individual patients in their clinical practice. Finally, if subjective eligibility criteria are a source of investigator uncertainty, the expense of a study may be increased and study enrollment slowed by longer eligibility assessment interviews and repeated calls from investigators with questions regarding eligibility to the clinical trial coordinating centre.

In this study, which involved cancer RCT investigators from across North America, we set out to answer two questions: Are subjective eligibility criteria associated with more variable enrollment decisions? Are investigators less certain of their decisions when using such criteria? The answers to these questions may have important implications for the design of future RCTs.

Methods

Over a period of months, our multidisciplinary research group developed the study instrument. Five common eligibility criteria, three subjective and two objective, were selected from an important series of NSABP breast cancer RCTs. The subjective criteria chosen were as follows:

- "[Exclude] patients who have nonmalignant systemic disease (cardiovascular, renal, hepatic etc.) which would preclude their being subjected to any of the treatment options (adjuvant chemotherapy) or prevent prolonged follow-up [criterion a, figures 1 and 2];"
- "Patients with psychiatric or addictive disorders which would preclude obtaining informed consent are ineligible [criterion c, figures 1 and 2];"
- "Patients must have a life expectancy of at least 10 years, excluding their diagnosis of cancer, to be enrolled in the study [criterion e, figures 1 and 2]."

In the set of NSABP breast cancer studies from 1972 to 1992 (NSABP B-04 to NSABP B-25), the criteria were present in 100%, 91% and 50% of the study protocols, respectively. The objective criteria were:

- "Patients with a previous or concomitant malignancy [are ineligible], regardless of site, EXCEPT patients with squamous or

basal cell carcinoma of the skin that has been effectively treated or carcinoma in situ of the cervix that has been treated by operation only [criterion b, figures 1 and 2];"

- "To be included in the study the patient must have a white blood cell count (WBC) greater than or equal to 4,000/mm^3 and a platelet count greater than or equal to 100,000/mm^3 [criterion d, figures 1 and 2]."

Again, the criteria were present in the majority of the NSABP studies: 100% and 86%, respectively.

For each eligibility criterion, questionnaire recipients were presented with clinical vignettes and asked whether they would enroll the patient in a breast cancer treatment protocol (yes or no) and to indicate on a visual analogue scale how sure they are of this decision (scored from 0.00 to 1.00). (Aitken, 1969; Folstein, 1973) Clearly, enrollment decisions will be a function, in part, of the clinical vignette. In order to control for this in the study design (and later in the analysis), three vignettes were given for each eligibility criterion: an eligible patient, an ineligible patient, and one whose status was uncertain. The order of the vignettes was varied for each eligibility criterion in the questionnaire. Recipients were asked if they had evaluated patients with any of the eligibility criteria in the questionnaire in their own experience with RCTs. Recipients were also asked for

demographic information, including medical specialty, years of medical practice, and number of RCTs in which they had participated. A pilot version of the questionnaire was given to 10 experienced clinical investigators in the Faculty of Medicine at McGill University to estimate the time required to complete the form, identify any problems in the construction of the instrument, and check the face-validity of the vignettes (i.e., face-validity as eligible, uncertain or ineligible). Seven investigators returned completed forms and, based on their comments, a three-page (plus instructions) final version of the questionnaire was prepared (see appendix 1).

Three-hundred-and-sixty-five oncologist investigators whose primary affiliation was with the U.S.-based NSABP or Canada-based National Cancer Institute of Canada (NCIC) were identified in the Physician Data Query database maintained by the National Library of Medicine (Bethesda, Maryland). After ethics approval for the study was obtained from the Research Ethics Board of the S.M.B.D. Jewish General Hospital (Montreal), each of the oncologist investigators in the sample was sent a card in October 1993 informing them of the study and that they would receive a questionnaire in two weeks time. The questionnaire, complete with instructions and a stamped and addressed return envelope, was mailed at the end of October 1993. With the first questionnaire mailing, recipients were told that they were free to not participate and refusal to participate should be

indicated by sending back a blank questionnaire. Investigators who failed to send back the questionnaire (either completed or blank) were sent up to two more mailings at four week intervals. The study was closed in February 1994.

Returned questionnaires were considered usable if the respondent filled in all of the primary outcome variables, i.e., the enrollment decisions for each of the vignettes. Data were entered into an electronic format, checked for transcription errors, and analysed using SAS. The primary outcome variable was the difference in enrollment decisions between subjective and objective criteria. Since we hypothesized that objective criteria would be associated with less variable enrollment decisions, i.e., decisions closer on average to one (enroll) or zero (not enroll), the difference had to be calculated in a different way for each of the three scenarios. Scenario-specific differences in enrollment decision were calculated in such a way that the difference would be positive if subjective criteria are associated with more variable enrollment decisions. Only if an analysis of variance for repeated measures (using PROC ANOVA in SAS) failed to conclude that the mean scenario-specific differences are different, would they be combined into a single summary measure. A t-test (using PROC MEANS in SAS) checks whether the differences (either scenario-specific or pooled) differ significantly from zero.

The secondary outcome is the certainty associated with each of the enrollment decisions. Since we hypothesized that, in each scenario, enrollment

decisions associated with subjective criteria would be less certain than those associated with objective criteria, the analysis is straightforward. For each scenario, the difference of interest was calculated as the mean certainty for the objective criteria minus the mean certainty for the subjective criteria. Calculated in this way, each of the differences would be positive if subjective criteria cause investigators to be more uncertain. As for the primary outcome analysis, the scenario-specific differences would only be combined if the results of the analysis of variance allow us. As above, t-tests will be used to test whether observed differences are significantly different from zero.

Finally, for both the primary and secondary outcome variables (whether scenario-specific differences or pooled differences are eventually used), multiple regression modelling (using PROC REG in SAS) explores the role of demographic factors in explaining the observed differences in enrollment decision or certainty.

Results

Of the 365 questionnaires sent out, 224 (61.4%) usable questionnaires were eventually returned (table 1). Perhaps due to the fact that our research group is based in Canada, the response rate was substantially higher for investigators associated with the NCIC RCT cooperative group (return rate = 78.1%, $p<0.00001$), although the U.S. returns compares favorably with similar

questionnaire surveys (response rate = 46.9%). Sixty-one percent of the respondents were medical oncologists, 24% were surgeons, and 14% were radiation oncologists (1% other). The surveyed investigators were both experienced physicians and clinical trialists: the investigators had practised medicine for an average of slightly more than 15 years and participated in roughly 28 RCTs. A large majority of the investigators (88.4%) had used at least one of the eligibility criteria listed in our questionnaire in their own experience with enrolling patients in RCTs.

Figure one shows the mean enrollment decision and 95% confidence interval for the five eligibility criteria under each of the three patient scenarios. If the scenario classification of the clinical vignettes is valid, we expect the mean enrollment decisions for the eligible scenario to be, on the whole, closer to one (i.e., enroll), for the ineligible scenario, closer to zero, and for the uncertain scenario, somewhere in between. The results confirm the face-validity of the patient scenarios.

Figure two shows the mean certainty associated with the enrollment decisions and 95% confidence interval for, once again, the five eligibility criteria under each of the three patient scenarios. As one might have expected, the mean certainty is less dependent on the patient scenario than mean enrollment decision. Overall, the enrollment decisions were associated with a fairly high degree of

certainty across the board, with all mean values for certainty exceeding 0.70.

The mean differences in enrollment decision for each patient scenario are listed in table 2. The analysis of variance indicates that the scenario-specific means are significantly different from one another ($F=213.51$, $df=2, 221$, $p=0.0001$) and, hence, it would be inappropriate to combine the differences into a pooled-difference measure. Ninety-five percent confidence intervals and p-values are, therefore, provided for each of the patient scenarios. In each patient scenario, the mean difference is positive, indicating that enrollment decisions for subjective criteria are more variable; the observed difference is marginally significant in the eligible-patient scenario ($p=0.07$), and highly statistically significant in the uncertain and ineligible-patient scenarios ($p=0.0001$ in each case). The magnitude of the observed difference is dependent on the patient scenario (as evidenced by the analysis of variance): the smallest difference was observed in the eligible-patient scenario, the intermediate difference in the ineligible-patient scenario, and, as one might have expected, the largest difference in the uncertain-patient scenario.

The mean differences in certainty associated with the enrollment decisions are also given in table 2. The analysis of variance indicates that the mean differences differ significantly from one another ($F=15.27$, $df=2, 221$, $p=0.0001$) and, hence, it would be inappropriate to combine them. For each of the patient

scenarios, the mean difference is positive and statistically significantly different from zero (p=0.0001 in each case), indicating that subjective criteria cause investigators to be more uncertain about enrollment decisions. The magnitude of the observed difference was dependent on the patient scenario, and the pattern mirrored that observed for the enrollment decisions: eligible-patient scenario, smallest difference; ineligible-patient scenario, intermediate difference; and, uncertain-patient scenario, largest difference.

The final stage of the analysis explored the explanatory value (if any) of the demographic variables collected on each respondent. Six separate multiple regression analyses using forward model selection were carried out: one for each of the two outcome measures under each of the three patient scenarios. In the analyses, the variables for years of medical practice and number of RCTs participated in were treated as continuous variables, clinical trial cooperative group affiliation was treated as a Bernoulli variable, and medical specialty was captured by a fixed group of three dummy variables. The results of the analysis are given in appendix 2. Clearly, in this sort of analysis multiple testing is an issue; we are, therefore, looking for obvious patterns across the patient scenarios. Of the twenty-four uncorrected p-values in the analysis, only two were statistically significant. (Without correction for multiple testing, we would expect one or two false-positive results with twenty-four tests). Each of the two "positive" variables was associated

with a very low partial r^2 (i.e., the variables explained a very small proportion of the variance in the outcome variable; roughly 4% in each case). Certainly, no strong pattern emerged from the analysis, and we concluded that none of the demographic variables explained observed differences in enrollment decisions or investigator certainty.

Discussion

The attitudes and behavior of RCT investigators has received some attention in the empirical literature. (Gotay, 1991; Hunninghake, 1987) For example, Taylor and colleagues examined the reasons physicians had for not enrolling otherwise eligible patients in an international breast cancer treatment study. (Taylor, 1984) Clinical vignettes have been used most commonly in studies using investigators as "expert surrogates" for the evaluation of the risk-benefit ratio of particular RCTs. (MacKillop, 1992) We believe that our study represents the first attempt to study enrollment decisions by investigators using clinical vignettes. Others may find our approach of use: investigators may wish to study other questions pertaining to physician behavior when enrolling patients in RCTs; indeed, RCT designers themselves may wish to do pilot studies of proposed eligibility criteria before including them in the final study protocol.

Our study hypothesized that subjective eligibility criteria would be

associated with more variable enrollment decisions and greater investigator uncertainty than objective criteria. The study findings support both hypotheses. In each of the three patient scenarios, enrollment decisions for subjective criteria were more variable than those for objective criteria (for the eligible-patient scenario, the difference was only of marginal statistical significance). The magnitude of the observed difference, however, depended on the patient scenario. Enrollment decisions for subjective criteria were associated with less investigator certainty than those for objective criteria in all three patient scenarios. The magnitude of the difference in certainty also varied with patient scenario, and the pattern of differences was similar to that observed for enrollment decisions. None of the observed differences were explained significantly by demographic characteristics of the investigators. It is important to note that most respondents were experienced physician-investigators, who had used these very same criteria in their clinical trials practice. Thus, the variability and uncertainty associated with subjective criteria cannot be explained by the respondents' unfamiliarity with using these concepts.

In the paper's introduction we pointed out that eligibility criteria associated with variable enrollment decisions and high investigator uncertainty may slow patient enrollment in the study, diminish the study's internal validity, and hamper the interpretation of study findings in clinical practice. The subjective eligibility

criteria that we chose for the study have been criticized as prejudicial. The results of our survey then provide yet another reason why criteria that exclude the elderly (either directly or indirectly), persons with psychiatric disease, or persons with a history of drug or alcohol abuse should be questioned by RCT designers, funding agencies, and Institutional Review Boards. Rather than aiding in resolving a scientific question, the presence of subjective criteria itself introduces an uncontrolled variable, of unknown significance, in delineating a trial's population. These criteria are therefore problematic on grounds of scientific validity as well as of ethics.

The degree of subjectivity implicated in any given criterion of trial eligibility is for these reasons problematic, and, other things being equal, by avoiding such criteria both ethical and scientific difficulties are avoided. It does not follow that subjective criteria can or should be avoided altogether. Consider the proposed criterion of exclusion: "subjects who, in the investigator's opinion, are unduly vulnerable to the risks associated with study participation." (Weijer, 1993) Although a more subjective criterion can scarcely be imagined, research ethics committees may feel constrained for moral and legal reasons to impose such a criterion as a condition for approval of some trials.

Subjectivity is only one dimension of eligibility criteria, and can be tolerated if a sufficiently compelling case is made. The need for sound judgment in

balancing such considerations implies that trial designers should provide reasons why they have chosen the specific eligibility criteria they employ, a measure that would aid in the evaluation of studies by peer review panels, granting agencies, and research ethics committees alike. (Freedman, 1994)

Table 1.

		CLINICAL TRIAL COOPERATIVE GROUP			
		NCIC	NSABP	P-value*	TOTAL
No. of questionnaires sent		169	196		365
No. of questionnaires returned and usable in analysis (and proportion of those sent out)		132 (0.781)	92 (0.469)	<0.0001	224 (0.614)
CHARACTERISTICS OF RESPONDENTS					
No. in each medical specialty (and proportion of respondents)				<0.0001	
	Radiation oncology	13 (0.098)	18 (0.200)		31 (0.138)
	Surgical oncology	20 (0.150)	33 (0.359)		53 (0.237)
	Medical oncology	99 (0.750)	38 (0.413)		137 (0.612)
	Other	0 (0.000)	3 (0.032)		3 (0.013)
Mean years of medical practice (and standard deviation [SD])		18.3 (7.8)	11.1 (9.0)	<0.0001	15.4 (9.03)
Mean no. of trials participated in (and [SD])		31.4 (30.6)	23.3 (29.1)	0.05	28.1 (30.2)
No. responding to the question, "Have you assessed patients for a clinical trial that has used one of the study eligibility criteria?" (And proportion of respondents)				0.006	
	Yes	124 (0.939)	74 (0.804)		198 (0.884)
	No	6 (0.046)	13 (0.141)		19 (0.085)
	Unsure	2 (0.015)	5 (0.054)		7 (0.031)

* - Unpaired t-test was used for continuous variables; Fisher's exact test was used for categorical variables.

Table 1. Response rates and characteristics of respondents.

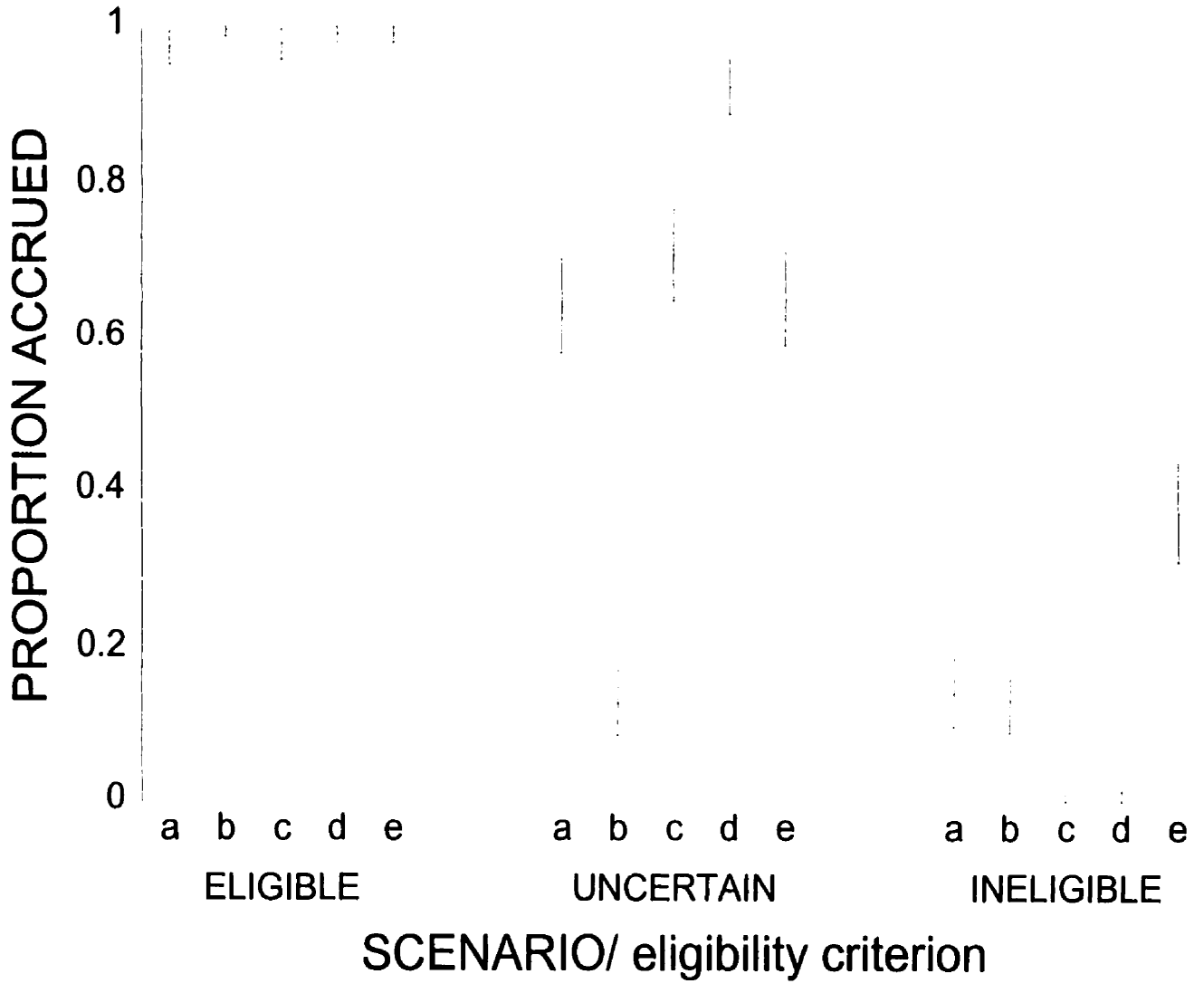

PROPORTION ACCRUED
1
0.8
0.6
0.4
0.2
0
a b c d e
a b c d e
a b c d e
ELIGIBLE
UNCERTAIN
INELIGIBLE
SCENARIO/ eligibility criterion

Figure 1. Accrual decisions for the clinical vignettes. For each eligibility criterion under each of the three patient scenarios, the mean proportion (and 95% confidence interval) of investigators who indicated they would enroll the patient in a trial is given. Eligibility criteria a,c, and e are subjective criteria; eligibility criteria b and d are objective criteria (see methods section for full text of criteria).

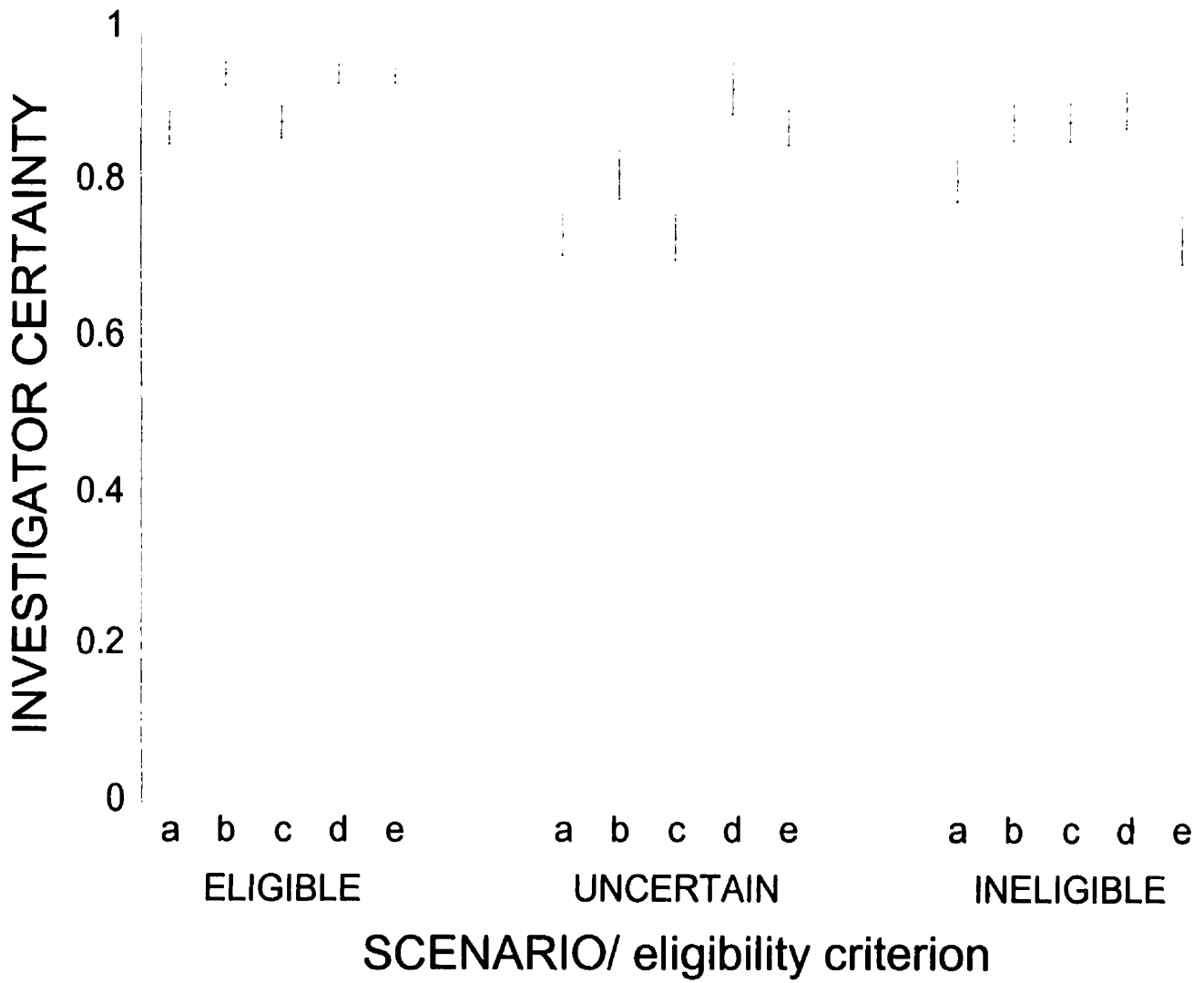

INVESTIGATOR CERTAINTY
1
0.8
0.6
0.4
0.2
0
a b c d e
a b c d e
a b c d e
ELIGIBLE
UNCERTAIN
INELIGIBLE
SCENARIO/ eligibility criterion

Figure 2. Investigator certainty associated with accrual decisions for the clinical vignettes. For each eligibility criterion under each of the three patient scenarios, the mean investigator certainty (and 95% confidence interval) associated with accrual decision is given. Eligibility criteria a,c, and e are subjective criteria; eligibility criteria b and d are objective criteria (see methods section for full text of criteria).

Table 2.

	PATIENT SCENARIO			ANOVA	
	ELIGIBLE	UNCERTAIN	INELIGIBLE	F	p-value
Mean difference in accrual decisions*	0.013	0.397	0.099	213.51	p=0.0001
95% confidence interval	-0.0009, 0.026	0.369, 0.425	0.062, 0.136		
p-value	p=0.07	p=0.0001	p=0.0001		
Mean difference in investigator certainty**	0.046	0.120	0.088	15.27	p=0.0001
95% confidence interval	0.035, 0.058	0.097, 0.142	0.065, 0.111		
p-value	p=0.0001	p=0.0001	p=0.0001		

* - positive values indicate that accrual decisions for subjective criteria are more variable than those for objective criteria.
** - positive values indicate that accrual decisions for subjective criteria are associated with less investigator certainty than those for objective criteria.

Table 2. Differences in accrual decisions and investigator uncertainty between objective and subjective eligibility criteria.

References

Aitken R.C.B. (1969) Measurement of feelings using visual analogue scales. *Proceeding of the Royal Society of Medicine* 62: 989-993.

Begg C. and Carbone P. (1983) Clinical trials and drug toxicity in the elderly; the experience of the Eastern Cooperative Oncology Group. *Cancer* 52; 1986-1992.

Begg C.B., Zelen M., Carbone PP, et al.. (1983) Cooperative groups and community hospitals: measurement of impact in the community hospitals. *Cancer* 52: 1760-1767.

Dresser R. (1992) Wanted single, white male for medical research. *Hastings Center Report* 22(1): 24-29.

Folstein M.F. and Luria R. (1973) Reliability, validity and clinical application of the visual analogue mood scale. *Psychological Medicine* 3: 479-486.

Freedman B. (1994) Multi-center trials and subject eligibility: should local IRBs play a role? *IRB: A Review Of Human Subjects Research* 16(1-2): 1-6.

Friedman M.A. and Cain D.F. (1990) National Cancer Institute sponsored comparative clinical trials. *Journal of the National Cancer Institute* 65: 2376-2382.

Fuks A., Weijer C., Freedman B., Shapiro S., Skrutkowska M. and Riaz A. (Unpublished data) A study in contrasts: eligibility criteria in a twenty-year

sample of NSABP and POG clinical trials.

Gotay C.C. (1991) Accrual to cancer clinical trials: directions from the research literature. *Social Science and Medicine* 33: 569-577).

Grisso T., Appelbaum P.S. (1991) Mentally ill and non-mentally-ill patients' abilities to understand informed consent disclosures for medication: preliminary data. *Law and Human Behavior* 15: 377-388.

Hughes J.R. (1993) Exclusion of "non-compliant" individuals from clinical trials. *Controlled Clinical Trials* 14: 176-177.

Hunninghake D.B., Darby C.A. and Probstfield J.L. (1987) Recruitment experience in clinical trials: literature summary and annotated bibliography. *Controlled Clinical Trials* 8: 6S-30S.

Kennedy B.J. (1991) Needed: clinical trials for older patients. *Journal of Clinical Oncology* 9: 718-720.

Kotwall C.A., Mahoney L.J., Myers R.E. and Decoste L. (1992) Reasons for nonentry in randomized clinical trials for breast cancer: a single institution study. *Journal of Surgical Oncology* 50: 125-129.

Lee J.Y. and Breaux S.R. (1983) Accrual of radiotherapy patients to clinical trials. *Cancer* 52: 1014-1016.

MacKillop W.J., Palmer M.J., O'Sullivan B. And Quirt C.F. (1992) The expert surrogate system. In: Williams C.J. (Ed.) *Introducing New Treatments for*

Cancer: Practical, Ethical and Legal Problems. John Wiley and Sons, New York.

Mastroianni A.C., Faden R. and Federman D. (1994) *Women and Health Research. Ethical and Legal Issues of Including Women in Clinical Studies. Volume I.* National Academy Press, 1994, Washington: 75-83.

McCusker J., Wax A. and Bennett J.M. (1982) Cancer patient accession into clinical trials: a pilot investigation into some patient and physician determinants of entry. *American Journal of Clinical Oncology* 5: 227-236.

Michels R. and Marzuk P.M. (1994) Progress in psychiatry (parts 1 and 2). *New England Journal of Medicine* 329: 552-560, 628-638.

Mor V., Masterson-Allen S., Goldberg R.J., Cummings F.J., Glicksman A.S., Fretwell M.D. (1985) Relationship between age at diagnosis and treatments received by cancer patients. *Journal of the American Geriatric Society* 33: 585-589.

National Institutes of Health. (1994) NIH guidelines on the inclusion of women and minorities as subjects in clinical research. *Federal Register* 59: 14508-14513.

Samet J., Hunt W.C., Key C., Humble C.G., Goodwin J.S. (1986) Choice of cancer therapy varies with age of patient. *JAMA* 255: 3385-3390.

Tate H.C. and Rawlinson J.B. (1979) Randomized comparative studies in the treatment of cancer in the United Kingdom; room for improvement? *Lancet* ii: 623-625.

Taylor K.M., Margolese R.G. and Soskilne C. (1984) Physicians' reasons for not entering eligible patients in a randomized clinical trial of surgery for breast cancer. *New England Journal of Medicine* 310: 1363-1367.

Weijer C. and Fuks A. (1993) The duty to exclude: excluding people at undue risk from research. *Clinical and Investigative Medicine* 17: 115-122.

Appendix 1. Study instrument.

INSTRUCTIONS:

The following eligibility criteria are taken from protocols for clinical trials in breast cancer. We would like you to apply them to scenarios that follow each of them. <u>For each scenario assume that the subject is otherwise eligible for the clinical trial</u>.

There are two questions for each scenario and we would like you to answer both of them. The first one asks if you would enrol the subject in light of the eligibility criterion and should be answered with a 'X' corresponding to your decision. The second one, has to do with how certain you are of your decision. You should answer the second question by making a single vertical line across the horizontal scale which corresponds to your certainty. (If you are "very unsure" or "very sure" do not circle the 'I' at the end of the line, rather, draw a line through it.) For example:

1. I would enrol this patient ( )
   I would not enrol this patient (X)
2. How certain are you of this decision?
   very unsure I----------/-------I very sure

Means that you feel that the patient is almost certainly ineligible.

Eligibility criterion: "[Exclude] patients who have nonmalignant systemic disease (cardiovascular, renal, hepatic etc.) which would preclude their being subjected to any of the treatment options (adjuvant chemotherapy) or prevent prolonged follow-up."

Scenario 1: A 58 year old woman with alcoholic cirrhosis of the liver and secondary esophageal varices. She has no ascites or episodes of hepatic encephalopathy. She is no longer drinking alcohol.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Scenario 2: A 50 year old woman who is an insulin dependent diabetic with recent onset peripheral neuropathy.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Scenario 3: A 60 year old woman with hypertension for 12 years, well controlled with medication (diastolic blood pressure on therapy approximately 85 mmHg).

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Eligibility criterion: "Patients with a previous or concomitant malignancy [are ineligible], regardless of site, EXCEPT patients with squamous or basal cell carcinoma of the skin that has been effectively treated or carcinoma in situ of the cervix that has been treated by operation only".

Scenario 1: A 51 year old woman who has had no previous malignancies. She has had three actinic keratoses removed from her face over the last 2 years.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Scenario 2: A fifty seven year old woman who is currently being investigated for a non- calcified pulmonary nodule in the upper lobe of the left lung. She has a twenty-five pack year history of smoking.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Scenario 3: A 69 year old woman who had an adenocarcinoma of the upper lobe of the right lung resected ten years ago followed by radiation therapy. She has been free of disease since that time.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Eligibility criterion: "Patients with psychiatric or addictive disorders which would preclude obtaining informed consent are ineligible."

Scenario 1: A 40 year old woman who had a single depressive episode five years ago which was effectively treated with ECT. She has had no subsequent symptoms.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Scenario 2: A 45 year old woman with a long history of paranoid schizophrenia, poorly controlled by medication.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Scenario 3: A 50 year old woman with bipolar affective disorder reasonably well controlled on lithium. There is no active symptomatology.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Eligibility criterion: "To be included in the study the patient must have a white blood cell count (WBC) greater than or equal to 4,000/mm^3 and a platelet count greater than or equal to 100,000/mm^3."

Scenario 1: A patient with a WBC of 4,500/mm^3 and platelet count of 130,000/mm^3.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Scenario 2: A patient with a WBC of 4,000/mm^3 and a platelet count of 100,000/mm^3.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Scenario 3: A patient with a WBC of 1,600/mm^3 and a platelet count of 110,000/mm^3.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Eligibility criterion: "Patients must have a life expectancy of at least 10 years, excluding their diagnosis of cancer, to be enrolled in the study."

Scenario 1: A 60 year old woman who had coronary angioplasty 18 months ago and is currently asymptomatic.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Scenario 2: A 62 year old woman with no significant past medical history.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Scenario 3: A 78 year old woman with no significant past medical history except for a twenty-two pack year smoking history.

1. I would enrol this patient ( )
   I would not enrol this patient ( )
2. How certain are you of this decision?

very unsure I----------I very sure

Please answer these questions:

1. What is your medical specialty? _____

2. How many years have you practised medicine? _____

3. Approximately how many clinical trials have you participated in ? _____

4. Have you ever assessed patients for a clinical trial that has used one of the above eligibility criteria ?

   yes ( )
   no ( )
   unsure ( )

5. Do you have any comments on the eligibility criteria given in this questionnaire? (use the back of this page if you need more space)

Appendix 2.

	PATIENT SCENARIO		
	ELIGIBLE	UNCERTAIN	INELIGIBLE
ACCRUAL DECISION			
Medical specialty			
order of entry into model	1	2	3
p-value at entry into model	NS**	NS	NS
proportion of variance explained*	0.020	0.005	0.029
Clinical trial cooperative group affiliation			
order of entry into model	2	1	2
p-value at entry into model	NS	NS	NS
proportion of variance explained	0.005	0.002	0.007
Years of medical practice			
order of entry into model	3	4	4
p-value at entry into model	NS	NS	NS
proportion of variance explained	0.009	0.0001	0.0007
No. of trials participated in			
order of entry into model	4	3	1
p-value at entry into model	NS	NS	p=0.0041
proportion of variance explained	0.002	0.0001	0.038
INVESTIGATOR CERTAINTY			
Medical specialty			
order of entry into model	2	1	1
p-value at entry into model	NS	P=0.0310	NS
proportion of variance explained	0.020	0.042	0.029
Clinical trial cooperative group affiliation			
order of entry into model	4	2	4
p-value at entry into model	NS	NS	NS
proportion of variance explained	0.0005	0.006	0.002
Years of medical practice			
order of entry into model	3	4	2
p-value at entry into model	NS	NS	NS
proportion of variance explained	0.004	0.0001	0.009
No. of trials participated in			
order of entry into model	1	3	3
p-value at entry into model	NS	NS	NS
proportion of variance explained	0.011	0.0008	0.003

* - partial r^2.

** - p-value greater than 0.05.

Appendix 2. Demographic characteristics of questionnaire respondents as potential explanatory factors for observed differences in accrual decisions and investigator uncertainty. (Multiple regression models using open forward selection procedure).

Acknowledgements

The study authors are grateful to Prof. Kathleen Cranley Glass, McGill University for her helpful comments on an earlier draft of the paper.

Bridging section

In the third chapter, my colleagues and I examined another step in the implementation of research subject selection procedures, namely, the interpretation of eligibility criteria by investigators. We hypothesized that subjectively-phrased criteria are a source of enrollment-decision variability and investigator uncertainty. If true, these factors may impact on the internal validity of a study, the study's efficient conduct, and even its eventual interpretation by clinicians in practice. We sought, therefore, to answer two questions: Are subjective eligibility criteria associated with more variable enrollment decisions? Are investigators less certain of their decisions when using such criteria?

In order to answer these questions, we developed a questionnaire which presents investigators with three hypothetical (but clinically relevant) patient scenarios for each of five common eligibility criteria, three subjective and two objective. For each scenario, investigators are asked whether they would enroll the patient in a study and how certain they are of this decision. 365 oncologist-investigators from the United States and Canada were invited to participate in the study, and 61.4% of those approached returned a usable questionnaire — a very good response rate for a study involving physicians.

We found that subjectively-phrased criteria were indeed associated with more variable enrollment decisions. The magnitude of the difference depended,

however, on the sort of patient the investigator was presented with: the effect was smallest when the patient was eligible, intermediate when the patient was ineligible, and largest when the patient's eligibility status was uncertain. Investigators were also more uncertain of enrollment decisions made for subjective eligibility criteria. Again, the magnitude of the effect varied with the patient scenario, as above.

Given the negative effects that subjective criteria may have on the validity, conduct and interpretation of clinical trial results, such criteria must be carefully justified. We suspect that in many, if not most, cases subjective criteria can be dropped in their entirety. In other cases, an important individual criterion may be rephrased in a more objective manner. A number of groups, including Institutional Review Boards, have an important role in reviewing eligibility criteria for their necessity.

In the next chapter, we turn from the implementation of eligibility criteria to their reporting in communications regarding study results. The faithfulness of reporting of criteria was identified at the end of chapter 2 as another important area for further empirical work. Thus, in chapter 4, we ask: Are eligibility criteria accurately reported in communications regarding the results of clinical trials? If reporting is not complete, what is the nature of information loss (what types of criteria are not reported)?

Chapter 4:

Reporting the study populations of clinical trials:

clear transmission or static on the line?

Abstract

Background. In contrast to attempts that have been made to measure the clarity of reporting of the methods of clinical trials in journal articles, we report here an attempt to measure the accuracy of methods reporting. As an exemplar of the reporting of clinical trial methods, we focus on eligibility criteria.

Methods. We examined the reporting of eligibility criteria in the protocol, methods paper (if applicable), journal article, and Clinical Alert for articles appearing in print between 1988 and September 1994 for which a Clinical Alert had been issued. Eligibility criteria were further classified into five categories in order to examine the content of information loss, if any.

Results. On average 82% of protocol eligibility criteria were reported in methods papers. Journal articles and Clinical Alerts fared somewhat worse: 63% of criteria were reported in journal articles; 19% in Clinical Alerts. In all three categories of medical communication, the reporting of criteria that defined the study disease tended to be complete; reporting of criteria relating to trial precision, patient safety, legal and ethical concerns, and administrative considerations, was not.

Conclusion. Criteria for clinical trial eligibility are frequently underreported in medical communications. Many of the criteria omitted were of clinical importance. Two courses of action are suggested: when clinical trials are designed, the need for individual eligibility criteria ought to be scrutinized; when trials are reported, care

should be taken to ensure that criteria are fully and accurately disclosed.

(238 words)

Keywords: clinical trials (methods), clinical trials (standards), eligibility determination, periodicals (standards).

Introduction

The game Broken Telephone is played with children seated in a circle. One child starts, whispering a message to the next child on the right; each child in turn passes along the whisper.[1] By the time it has come full circle, the message is distorted, sometimes beyond recognition. Medical information, too, gets channelled through a variety of sources before reaching the ears of medical practitioners, who will interpret the information and attempt to apply it to the care of their patients. Just as in the children's game, the information may become degraded in a stepwise fashion. In this article, we discuss and demonstrate this effect with respect to an important aspect of clinical trials: the definition of the study population that emerges from a trial's criteria of eligibility.

The primary methodological advantage of the randomized clinical trial lies in the accuracy it affords. It provides a mechanism for control of selection bias and potentially confounding variables via the use of randomization and the advance definition of the trial's parameters. Trial designers typically go to great lengths to precisely define the population and the condition under study. In determining subject eligibility, special attention is given to characteristics of the disease (for example, in early stage breast cancer trials, estrogen and progesterone receptors are quantified), as well as of the patient (for example, excluding patients with intercurrent illnesses, those refractory to previous treatment, etc.) Unless these trial

characteristics are clearly defined and rigorously applied, neither the relevance of the trial's results for resolving the question it had originally posed (internal validity), nor the trial's implications for treatment recommendations (external validity or generalizability), may be relied upon.

Proper design and conduct of a trial is not, however, sufficient. Its methods and results need to be accurately and comprehensively communicated. From a scientific point of view, such reporting is necessary for a study's proper evaluation and, potentially, replication or refutation. From a clinical perspective, gaps and errors in the conduct and reporting of a trial may directly lead to inappropriate treatment decisions. From an ethical point of view, moreover, society expects trialists to satisfy the highest standards of candour and accuracy, as has been codified, for example, within the World Medical Association's Helsinki declaration on the ethics of human research.[2]

Recently, a great deal of attention has focused on scientific fraud, the deliberate misrepresentation and/or falsification of data. Reports of putative misconduct are dramatic, and quickly capture the attention of the professional and lay community.[3] A less sensational concern, but one which very probably poses a greater challenge to the scientific literature, is the underreporting of negative research findings, and the attendant bias that this can produce in the published literature.[4,5,6]

Several sources assert that there exists an obligation to accurately report all aspects of a trial, including the eligibility criteria proposed (and actually employed).[2,7,8,9] The authoritative Uniform Requirements for Manuscripts Submitted to Biomedical Journals instructs authors to "[d]escribe your selection of observational or experimental subjects...clearly".[7] In reporting the subject selection process, sufficient detail must be provided to allow "other workers to reproduce the results".[7] In their published commentary on the Uniform Requirements (1988 edition), Bailar and Mosteller[10] describe two basic goals that mandate a precise description of subject selection: first, to allow other investigators attempting to replicate the study to "make nearly the same decisions about including patients in the study"; and, second, to provide readers interpreting the published report "with a solid link between the patients or cases studied and the population for which the inferences will be made". But do published reports of clinical trials live up to this requirement?

Chalmers et al.[11] developed an instrument to measure the quality of clinical trials which includes an assessment of the adequacy of the description of subject selection procedures. This research tool has subsequently been used by a number of authors to measure the quality of published reports of trials. Zola et al. used Chalmer's methodology in a review of 152 published reports on the treatment of early cervical cancer. They found that information on patient characteristics was

adequately reported in only 7% of the studies.[12] Liberati et al. studied published reports of clinical trials of the treatment of early stage breast cancer.[13] They found that the description of patient selection criteria was sufficient in only 46% of the studies.

DerSimonian et al.[14] devised an 11-point instrument specifically for rating published reports of clinical trials. As in Chalmers' research tool, one aspect examines the adequacy of disclosure of criteria for clinical trial eligibility. In their review of 67 clinical trials reported in 4 major medical journals, DerSimonian et al. reported that only 37% of articles adequately reported study criteria.[14] Emerson et al.[15], using DerSimonian's method, found that trials reported in surgical journals followed the same pattern: only 43% of articles reported selection criteria adequately. It seems clear from these published studies that, despite the obligation to report the study selection process of a trial fully, many trial reports fall short of meeting the requirement.

The focus of prior studies has taken the published journal article as the unit of investigation. For example, DerSimonian et al.[14] examined the quality of reporting of trial methods by a subjective assessment of the clarity of the reporting. A journal article is not however a self-contained text, to be judged by literary standards (such as clarity) alone. To measure accuracy of reporting, we need to compare the contents of journal articles with the original and controlling scientific

protocol.

For that matter, articles in general medical journals are not the only means that exist for communicating the results of clinical trials. A practitioner may attend more closely to a brief clinical recommendation that has been issued following a trial's conclusion; a researcher may instead focus upon the methodological description of a trial that appears in a more specialized literature. In this study, therefore, we have directly compared the eligibility criteria that are found in the study's protocol with those criteria that appear in a series of subsequent communications: the trial's methodology paper (hereafter, the 'methods paper'); the trial's final report (hereafter, the 'journal article'); and the advisories issued by the National Institutes of Health (hereafter, the 'Clinical Alert') under their Clinical Alert system.

The Clinical Alert system provided a useful focus for this initial examination of the problem of dissemination of information about trials. Since their inception in 1988, Clinical Alerts, have been issued by NIH to notify the medical community of important new information gathered from clinical trials.[16,17] The trials that have resulted in Clinical Alerts are therefore an elite group that meet high standards of scientific rigor and clinical relevance. Reports of these trials appear in the leading medical journals. Thus by studying trials that have resulted in the NIH issuance of a Clinical Alert, we may focus more directly on the issue of

dissemination itself: The problems we describe below cannot plausibly be said to be the artifactual result of slipshod trial design or substandard practices of medical publication. Finally, few trials have been seen by NIH to be of such immediate importance as to merit a Clinical Alert; the set of these studies therefore yields a manageable number for investigation.

Methods

We selected for study all of the trials that served as the subject of a single study Clinical Alert for which the journal article was published between 1988 and September of 1994. (Our methodology precluded the inclusion of Clinical Alerts that were based upon the results of several studies, such as the first Clinical Alert regarding the treatment of node-negative breast cancer.[18]) We then collected the corresponding methods paper, if available, and journal article. The corresponding author listed on the final journal article was contacted for a full-text copy of the clinical trial protocol.

For each clinical trial, eligibility criteria in the protocol, methods paper (if available), journal article and clinical alert were counted by at least two of the three authors. Differences were discussed at the time of counting and reconciled through discussion and group consensus. In counting the criteria, an attempt was made to minimize artifactual difference between the different reports of any

specific trial. For example, the exclusion criterion "patients with severe heart disease" reported in a journal article was counted as three criteria in that it replaced the exclusion criteria "patients with severe cardiac valvular disease", "patients with an MI within the last six months" and "patients with angina requiring chronic medication" present in the protocol. We have standardized our results by expressing them as a proportion of criteria reported in the original clinical trial protocol.

We then attempted to characterize qualitative aspects of the information loss by classifying each of the clinical trial protocol eligibility criteria according to a schema developed by us previously.[19] In brief, the schema partitions eligibility criteria into five mutually-exclusive categories (all examples are taken from the Studies of Left Ventricular Dysfunction (SOLVD) Prevention and Treatment Trials protocol):

- Definition of disease -- criteria that define clinical parameters of the disease being studied, e.g., "left ventricular ejection fraction of less than or equal to 35%";
- Precision -- criteria that render the study population more homogeneous for the purposes of the trial, e.g., "[exclude patients with] malignancies, except for surgically cured skin cancer, carcinoma-in-situ, or 5 years free of disease";
- Safety -- criteria that exclude persons thought to be unduly vulnerable to harm from the study therapy, e.g., "[patients with] cerebrovascular disease...that could

potentially be complicated or rendered unstable by the administration of an ACE inhibitor [are excluded]";

- Legal and ethical -- criteria needed to ensure that research satisfies legal and ethical norms of human experimentation, e.g., "failure to give consent" (exclusion criterion); and,
- Administrative -- criteria that ensure the smooth functioning of the trial, e.g., "likelihood of prospective participant being nonadherent due to chronic alcoholism..." (exclusion criterion).

As before, counts in each category were expressed as a proportion of the number of criteria per category in the trial protocol. The number of trials available for study is small, and we present our results descriptively, without formal attempts at inference.

Results

In all, eight single-trial Clinical Alerts were released by the NIH for trials with corresponding journal articles appearing between 1988 and September of 1994. We were successful in obtaining protocols for all eight clinical trials corresponding to the Alerts.[20,21,22,23,24,25,26,27] A methods paper was published for 5 of the trials[28,29,30,31,32] and a final report was published for each of the eight studies.[33,34,35,36,37,38,39,40] The clinical trials included in our analysis are listed in

Table 1.

Tracking down the journal article for each Clinical Alert, as a practitioner who received the Alert might have attempted to do, proved to be far from a simple or speedy task. Obtaining the journal article was hindered by two factors. First, as indicated in Table 1, many of the Clinical Alerts were published substantially in advance of the journal article. Indeed, the median Alert-to-article publication delay was four months (120 days). Second, only three of the seven Clinical Alerts provided a reference (journal and issue) for the final-report articles.

Counts and proportions of eligibility criteria reported for each of the clinical trials are detailed in Table 1 and graphically displayed in Figure 1. Counts and proportions for each category of eligibility criteria are listed in Table 2. (We have not listed the precise counts by category for each individual trial in the interest of conserving space.)

Overall the reporting of eligibility criteria in the methods papers was quite complete ($\pi = 0.82$)(Figure 1). The information loss in the methods papers was characterized by an underreporting of those criteria dealing with administrative aspects of the trial (Table 2). As these criteria are directed at such issues as subject proximity to the study centre, they are not key to the generalizability assessment made by practitioners.

The reporting of eligibility criteria in journal articles was, however, quite

incomplete. On average, only 63% of protocol eligibility criteria were reported in the corresponding journal article. Part of the eligibility criteria information loss in journals may be due to a reliance on the disclosure in methods papers. Indeed, referring to Figure 1, 4/4 of the lowest reporting journal articles were preceded by methods papers whereas only 1/4 of the highest reporting articles were.

Few criteria addressing 'legal and ethical' and 'administrative' aspects were reported, but these are of little relevance to the practising clinician. Of greater concern, only 66% of 'precision criteria' were reported. In that precision criteria restrict the study population by excluding, for example, patients with concomitant medical conditions, failure to report them can interfere with the clinician's ability to assess the applicability of the trial to her own patients.

Then too, the fact that only 57% of the safety criteria were reported is disconcerting. The failure to report which subgroups of patients have been excluded out of concern for undue toxicity may lead the practitioner to believe that an intervention is safe in these particular subgroups. Certainly, the inclusion of a potentially vulnerable subgroup in a "positive" trial does not ipso facto demonstrate the safety of the study treatment in that subgroup. Nonetheless, if trial designers were sufficiently concerned about undue toxicity in a given subgroup of patients as to choose to exclude them from the trial, this concern ought to be relayed to practitioners.

The reporting of eligibility criteria was poorest in the Clinical Alerts. On average, only 19% of the criteria present in the protocol were disclosed in the Clinical Alert. Qualitatively, the pattern of information loss corresponded to that seen in the journal articles (Table 2). Despite the scant information presented in Alerts, only 2/8 of the Clinical Alerts included any statements which cautioned physicians regarding the interpretation of the clinical trial's results.

The omissions of eligibility criteria in the Clinical Alerts spanned the spectrum from trivial to serious. Most of the Alerts failed to mention that informed consent was obtained from research subjects. Consent itself is, of course, important; failure to disclose that it had been sought and obtained, less so. However, some Clinical Alerts, more seriously, failed to mention groups that were excluded for safety reasons. For example, the Clinical Alert for the Secondary Prophylaxis of Pneumocystis Carinii Pneumonia trial failed to mention the exclusion of patients with abnormal liver function, abnormal renal function, leukopenia or thrombocytopenia.[23] Some Alerts failed to report that a new treatment had not been evaluated in a subset of patients. For example, the Clinical Alert for the Studies of Left Ventricular Dysfunction (SOLVD) trial failed to mention that patients with serious coronary artery disease had been excluded from the trial and thus that the benefit of enalapril for this subgroup of patients remained unproven.[25] The most disconcerting case involved the Clinical Alert reporting the

results of the Optic Neuritis Treatment Trial. This trial found oral prednisone to be ineffective in the treatment of optic neuritis.[26] The Alert failed to mention, however, that the study was restricted to cases of idiopathic optic neuritis. The reader of the Alert may not realize that this finding is inapplicable to the approximately 50% of neuritis patients who suffer from its non-idiopathic form.[41]

It seems clear from the literature on Clinical Alerts that they provide physicians with information in a manner which invites them to consider altering their practice.[42,43] It is telling that the publication of the first Clinical Alert[18] (which, as a multi-trial Alert, was not included in our sample) caused 75% of clinicians surveyed to change their practice as a result.[44,45] Given that Clinical Alerts seem to exert a substantial influence on medical practice, they should present a more complete picture of the study population.

Discussion

A crucial factor in the improvement of medical care derives from the transfer of research to clinical practice. Clinical trials represent an essential step in this process. The number of patients enrolled in any single trial or even a collection of related trials usually represents a small proportion of the prospective patient population for the treatment under study. A pivotal issue in the successful diffusion of findings is not the size of this proportion, but its representativeness

and relevance. The complete and accurate reporting of eligibility criteria is an essential prerequisite to physician assessment of the generalizability of the results of a clinical trial.

Progress in medical knowledge commonly occurs in a slow and measured fashion. Advances build upon the successes and failures of previous efforts in a way that underlines the etymology of the word 're-search' -- *rechercher*, to search again. Given this incremental pattern of progress, full and accurate reporting of both the methodology and results of previous research is necessary.

We had chosen to study eligibility criteria as a test case for dissemination. In their review of the reporting of clinical trial results, DerSimonian et al. found eligibility criteria to be amongst the poorest reported categories of trial information.[14] Other researchers have suggested that eligibility criteria escape the kind of scrutiny that is expected of such an important component of clinical trials, noting very substantial disparity between the criteria employed by different trial groups working on parallel investigations (adjuvant chemotherapy for breast cancer).[46] It seemed to us probable that if inadequate dissemination were to be found regarding any aspect of clinical trials, it would be found in eligibility criteria.

Further study is needed concerning the dissemination of other important elements of trials. What we have shown about eligibility criteria is however

enough to state that there exists a serious gap in scientific communication, that can result in inappropriate extrapolation (or, inappropriate failure of extrapolation) of trial results to clinical practice. Using our findings as a test case, Glass has argued that this failure could ground legal liability, on the part of reporting investigators or editors of journals.[47] However, even without the spur of potential liability the biomedical research and publication community should act to rectify this communicative distortion.

What kind of responses should be considered, and on whose part? The following is a partial list:

•The Clinical Alert system has been criticized for prematurely extracting advice for clinical practice from just-completed clinical trials; in some cases, it was claimed, the trial results had not yet undergone an adequate peer-review process.[48] These charges have elicited a spirited defence of the system.[16,17] Underlying the empirical dispute over whether the Clinical Alerts issued reflected a complete and mature scientific consensus, there appears as well to be some underlying philosophical and policy dispute as to who should point out the clinical relevance of medical trials, and what professional and peer mechanisms should be utilized towards that end.

Our own work has no direct bearing upon this empirical controversy. Philosophically, we are sympathetic to the underlying effort on the part of the

authors of the Clinical Alert system to bring scientific results to the attention of clinicians in a prompt manner, reducing the distance between the 'bench' and the 'bedside'. For that to be done responsibly, however, the Alert must contain more information about the trial that serves as its basis than did those Alerts we had studied. A clinician acting in good faith may be said to be in an impossible position: On the one hand, decisions about patient care should reflect the kind of up-to-date information the Alert affords; on the other hand, knowing that important details concerning the population have not been revealed, the prudent practitioner must wait until he or she can determine whether a new treatment seems to be indicated for this patient. The cautious practitioner may feel the need to await the more complete presentation to be found in medical journals, but this of course defeats the original purpose of the Alert system, as this delay will on average amount to 4 months (as noted above).

•Journal publication raises some knotty problems. As discussed, methods papers do a relatively good job at describing eligibility criteria, but the main results papers appearing in journals of general medical interest are incomplete in this respect. We are not in a position to address the etiology of the problem. Authors may be choosing to not fully report the criteria that have been used in the trial, or reviewers and editors, as the guardians of readers' interests, may be choosing to filter criteria, presumably to avoid having a report's message obscured by what is

judged to be excessive detail. In either case, the decision may be motivated by the belief that the concerned reader of the medical journal can always turn to the methods paper for a more complete account.

There are however several problems with this approach. First, even within our small and highly select sample, not all studies produced a methods paper; we would expect that a more representative sample of the published medical literature would have been substantially less likely to have a parallel methodology publication. Within our sample, those studies that had no accompanying methods paper did a relatively good job at describing their eligibility criteria in their results paper; it remains to be seen whether this would continue to hold true for studies with a lower profile than ours. Second, not all of the readers of the general medical literature will have ready access to the methods papers that are published. This is particularly true of the isolated practitioner, but it is arguable how often any practitioner without an active academic interest in a particular disease accesses the methodology literature. In numerous discussions with colleagues about our interest in the question of dissemination, we often encountered a general, albeit unexamined, belief that a medical journal's report of a clinical trial is basically self-contained, at least as concerns eligibility issues. Third, to allow methods papers to fill the gap left by a results paper, the latter must be flagging that fact, by a comment to the effect that "A complete listing of eligibility criteria can be found

in ...", the responsibility for complete reporting had been appropriately discharged. However, none of the reports we reviewed for this study contained such an indication.

●Eligibility criteria themselves deserve further attention, on the part of trial designers and granting agencies. The rationale for each eligibility criterion used in a trial is usually not available to the practitioner deciding whether to apply the treatment in question to a particular patient. Often the rationale is not even available in the trial protocol. Although the rationale for some criteria may be self-evident, there are usually many criteria whose motivation is not transparent and requires documentation. The process of doing so may be illuminating at times and trial designers may find that certain proposed criteria do not have a very compelling underlying rationale.

What we have found in our study speaks to an even more basic point. Whether or not the rationale for eligibility criteria are presented, the criteria themselves are often not communicated as part of the results obtained. It would seem that either these criteria were important to the conduct of the trial-- in which case they must be communicated; or, they were of at most marginal importance -- in which case, trialists may be better advised to leave them out, resulting in a simpler trial, and one more true to the clinical reality of the patient population.[49]

On a broader scale, the results of our study suggest the need to reexamine

the concept and mechanics of the ethical review of the conduct of research, in particular the issues of its scope and timing. At present, research review is done in advance of its performance. A committee (IRB) examines the research plans, including the protocol and suggested consent forms. These plans are evaluated according to the common ethical principles of respect for persons, beneficence, and justice.[50]

This model of prior approval assumes that for the most part, the important ethical issues posed by research arise, and can be resolved, before the research has begun -- even before the first patient has been enrolled. By contrast, we suggest that research be viewed not as a proposal, but as a process, that exists within a context of scientific investigation and clinical progress. Ethical issues of a study can arise at any point throughout its conception, conduct and communication,[51] and it would therefore be strange to confine organized ethical evaluation, on the part of an investigator or of an institution within which a study is conducted, to a study's prologue.

Although the current attention to the institutional duty to monitor the conduct of research is a step in this direction,[52] a compelling case can be made that the journey is a longer one. There is a duty on the part of scientists to communicate the results they have obtained honestly, accurately, and completely. We do not believe that any of the participants in the dissemination of research dispute the

importance of this obligation to the scientific and clinical communities, any more than they dispute the obligation to treat the subjects of their research with respect (the principle of respect for persons) and with concern for their well-being (beneficence).

Table 1.

Study	Clinical trial	Alert-journal interval	Reporting of eligibility criteria			
		days	Protocol	Methods paper	Journal article	Clinical Alert
			number		proportion	
1	North American Symptomatic Carotid Endarterectomy Trial[20,28,33]	171	25	0.96	0.88	0.52
2	Efficacy of IVIG in Treatment of Symptomatic HIV-infected Children[21,34]	176	13		1.00	0.38
3	CMV Retinitis Trial[22,29,35]	69	17	0.82	0.35	0.18
4	Secondary Prophylaxis of Pneumocystis Carinii Pneumonia[23,36]	430	20		0.85	0.15
5	ddI versus ddC in HIV-infected Patients Who Are Intolerant of or Who Have Failed ZDV Therapy[24,37]	402	22		0.95	0.09
6	Studies of Left Ventricular Dysfunction Prevention and Treatment[25,30,38]	1	32	1.00	0.41	0.09
7	Systolic Hypertension in the Elderly Program[26,31,39]	0	22	0.55	0.32	0.09
8	Optic Neuritis Treatment Trial[27,32,40]	37	31	0.77	0.29	0.03
	Mean	(Median = 120)		0.82	0.63	0.19
	(Standard deviation)			(0.18)	(0.31)	(0.17)

Table 1. Reporting of eligibility criteria in clinical trial communications.

Table 2.

Category of eligibility criteria	Protocol	Methods paper	Journal article	Clinical Alert
	Number		Proportion	
Definition of disease	22	1.00	1.00	0.71
Precision	95	0.86	0.66	0.17
Safety	42	0.80	0.57	0.03
Legal and ethical	15	0.83	0.52	0.14
Administrative	9	0.20	0.17	0.00

Table 2. Reporting of eligibility criteria in clinical trial communications by category of eligibility criteria.

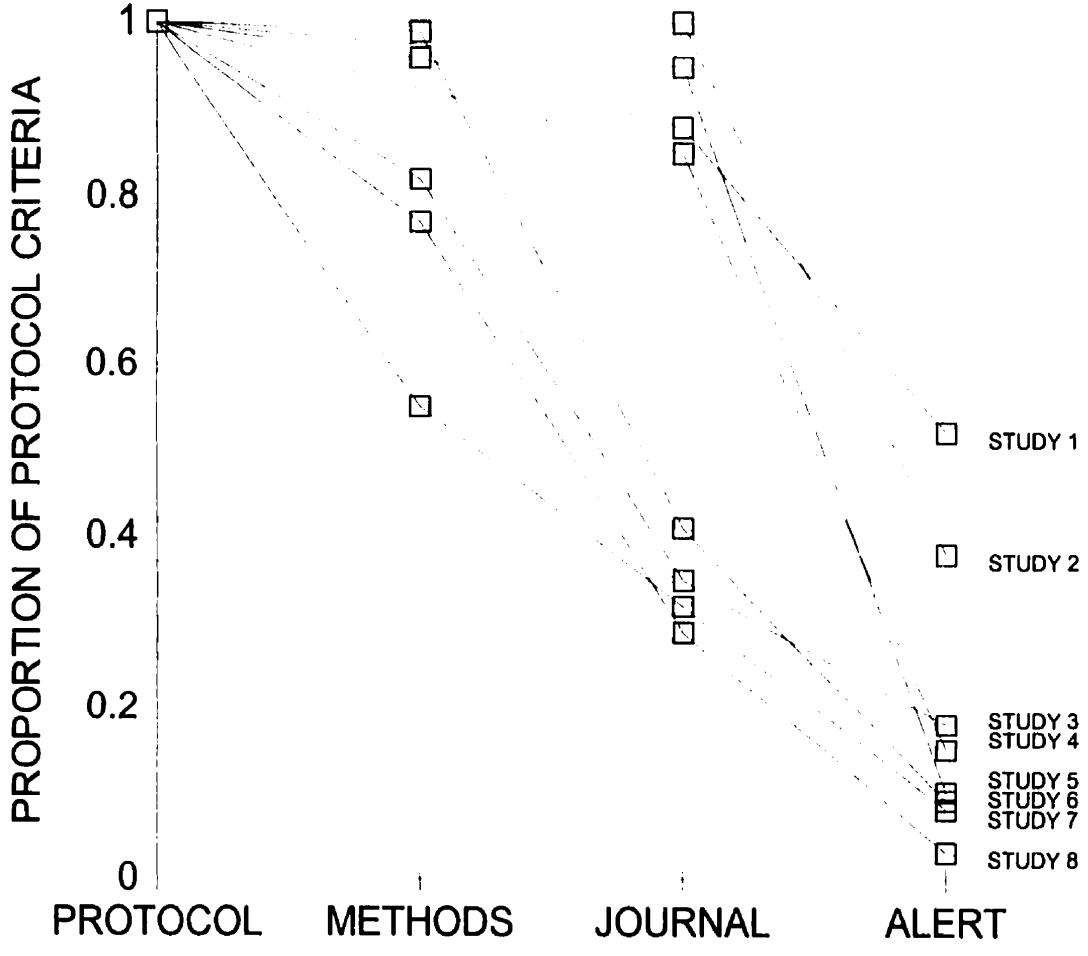

PROPORTION OF PROTOCOL CRITERIA
1
0.8
0.6
0.4
0.2
0
PROTOCOL
METHODS
JOURNAL
ALERT
STUDY 1
STUDY 2
STUDY 3
STUDY 4
STUDY 5
STUDY 6
STUDY 7
STUDY 8

Figure 1. Proportion of protocol criteria appearing in clinical trial communications.

References

1.Luetke-Stahlman B. Communication games. Perspectives for Teachers of the Hearing Impaired 1987; 5(3): 9-11.

2. World Medical Association Declaration of Helsinki: Recommendations Guiding Medical Doctors in Biomedical Research Involving Human Subjects. Adopted by the 18th World Medical Assembly, Helsinki, Finland, 1964, and as revised by the 29th World Medical Assembly, Tokyo, Japan, 1975.

3.Dingell J D. Misconduct in medical research. N Engl J Med 1993; 328: 1610-5.

4.Dickerson K. The existence of publication bias and risk factors for its occurrence. JAMA 1990; 263: 1385-89.

5.Chalmers I. Underreporting research is scientific misconduct. JAMA 1990; 263: 1405-8.

6.Dickerson K, Berlin JA. Meta-analysis: state-of-the science. Epidemiol Rev 1992; 14: 154-76.

7.International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. N Engl J Med 1991; 324: 424-428.

8.Simon R, Wittes RE. Methodologic guidelines for reports of clinical trials. Cancer Treat Rep 1985; 69: 1-3.

9.Zelen M. Guidelines for publishing papers on cancer clinical trials: responsibilities of editors and authors. J Clin Oncol 1983; 1: 164-169.

10.Bailar JC, Mosteller F. Guidelines for statistical reporting in articles for medical journals. Ann Intern Med 1988; 108: 266-273.

11.Chalmers TC, Smith H, Blackburn B, et al.. A method for assessing the quality of a randomized control trial. Control Clin Trials 1981; 2: 31-49.

12.Zola P, Volpe T, Castelli G, et al.. Is the published literature a reliable guide for deciding between alternative treatments for patients with early cervical cancer? Int J Radiation Oncology Biol Phys 1989; 16: 785-797.

13.Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized controlled trials of primary treatment of breast cancer. J Clin Oncol 1986; 4: 942-951.

14.DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. N Engl J Med 1982; 306: 1332-1337.

15.Emerson JD, McPeek B, Mosteller F. Reporting clinical trials in general surgical journals. Surgery 1984; 95: 572-579.

16.Friedman MA. The clinical announcement policy of the National Cancer Institute. In: Williams CJ (ed.). Introducing New Treatments for Cancer: Practical, Ethical and Legal Problems. New York: John Wiley & Sons, 1992: 413-419.

17.Wittes RE. Of Clinical Alerts and peer review. J Natl Cancer Inst 1988; 80: 984-985.

18.National Cancer Institute. Clinical Alert. Bethesda, Md.: National Cancer

Institute. May 16, 1988.

19.Fuks A, Weijer C, Freedman B, Shapiro S, Skrutkowska M, Riaz A. Eligibility criteria in breast cancer clinical trials. Manuscript in preparation.

20.National Institutes of Health. Clinical Alert. Bethesda, Md.: National Institutes of Health. February 25, 1991.

21.National Institutes of Health. Clinical Alert. Bethesda, Md.: National Institutes of Health. January 16, 1991.

22.National Institutes of Health. Clinical Alert. Bethesda, Md.: National Institutes of Health. November, 1991.

23.National Institutes of Health. Clinical Alert. Bethesda, Md.: National Institutes of Health. October 21, 1991.

24.National Institutes of Health. Clinical Alert. Bethesda, Md.: National Institutes of Health, February 1, 1993.

25.National Institutes of Health. Clinical Alert. Bethesda, Md.: National Institutes of Health. July 31, 1991.

26.National Institutes of Health. Clinical Alert. Bethesda, Md.: National Institutes of Health. June 26, 1991.

27.National Institutes of Health. Clinical Alert. Bethesda, Md.: National Institutes of Health. January 21, 1992.

28.North American Symptomatic Carotid Endarterectomy Trial Steering Committee. North American Symptomatic Carotid Endarterectomy Trial: Methods, patient characteristics and progress. Stroke 1991; 22: 711-720.

29.Studies of Ocular Complications of AIDS (SOCA) Research Group in collaboration with the AIDS Clinical Trials Group (ACTG). Studies of Ocular Complications of AIDS Foscarnet-Gancyclovir Cytomegalovirus Retinitis Trial: 1. Rationale, design, and methods. Control Clin Trials 1992; 13: 22-39.

30.The SOLVD Investigators. Studies of Left Ventricular Dysfunction (SOLVD) -- Rationale, design and methods: Two trials that evaluate the effect of enalapril in patients with reduced ejection fraction. Am J Cardiol 1990; 66: 315-322.

31.The Systolic Hypertension in the Elderly Program (SHEP) Cooperative Research Group. Rationale and design of a randomized clinical trial on prevention of stroke in isolated systolic hypertension. J Clin Epidemiol 1988; 41: 1197-1208.

32.Cleary PA, Beck RW, Anderson MM, et al.. Design, methods and conduct of the Optic Neuritis Treatment Trial. Control Clin Trials 1993; 14: 123-142.

33.North American Symptomatic Carotid Endarterectomy Trial Collaborators. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. N Engl J Med 1991; 325: 445-453.

34.The National Institute of Child Health and Human Development Intravenous Immunoglobulin Study Group. Intravenous Immune Globulin for the prevention of

bacterial infections in children with symptomatic human immunodeficiency virus infection. N Engl J Med 1991; 325: 73-80.

35.Studies of Ocular Complications of AIDS Research Group in collaboration with the AIDS Clinical Trials Group. Mortality in patients with the Acquired Immunodeficiency Syndrome treated with either foscarnet or gancyclovir for cytomegalovirus retinitis. N Engl J Med 1992; 326: 213-220.

36.Hardy WD, Feinberg J, Finkelstein DM, et al.. A Controlled Trial of Trimethoprim-Sulfamethoxazole or Aerosolized Pentamidine for Secondary Prophylaxis of Pneumocystis Carinii Pneumonia in Patients with the Acquired Immunodeficiency Syndrome. N Engl J Med 1992; 327: 1842-1848.

37.Abrams DI, Goldman AI, Launer C, et al.. A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection. N Engl J Med 1994; 330: 657-662.

38.The SOLVD Investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. N Engl J Med 1991; 325: 293-302.

39.SHEP Cooperative Research Group. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension: Final results of the Systolic Hypertension in the Elderly Program (SHEP). JAMA 1991; 265: 3255-3264.

40.Beck RW, Cleary PA, Anderson MM, et al.. A randomized, controlled trial of corticosteroids in the treatment of acute optic neuritis. N Engl J Med 1992; 326: 581-588.

41.Lessell S. Corticosteroid treatment of acute optic neuritis. N Engl J Med 1992; 326: 634-5.

42.Macdonald JS. Sometimes a great notion. J Natl Cancer Inst 1990; 82: 102, 104.

43.Wittenberg CK. Clinical Alerts: Broder discusses pros and cons on record. J Natl Cancer Inst 1990; 82: 186-187.

44.DeVita VT. Is a mechanism such as the NCI's Clinical Alert ever an appropriate alternative to journal peer review? Important Adv Oncol 1991: 241-253.

45.Clinical Alert stimulates changes in practice. J Natl Cancer Inst 1988; 80: 1185-1187.

46.Begg CB, Engstrom PF. Eligibility and extrapolation in cancer clinical trials. J Clin Oncol 1987; 5: 962-968.

47.Glass KG. Towards a duty to report clinical trials accurately: the Clinical Alert and beyond. J Law Med Ethics 1994; 22: 327-338.

48.Hellman S. Clinical Alert: a poor idea prematurely used. Important Adv Oncol 1991: 255-257.

49.Yusuf S, Collins R, Peto R. Why do we need large, simple randomized trials? Stat Med 1984; 3: 409-420.

50.U.S. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Washington, D.C.: U.S. Government Printing Office, 1978. DHEW Publication No. (OS) 78-0012. Reprinted in Federal Register 44 (April 18, 1979): 23192.

51.Freedman B, Shapiro S. Ethics and statistics in clinical research: towards a more comprehensive examination. Journal of Statistical Planning and Inference 1994; 42: 223-240.

52.Weijer C, Shapiro S, Fuks A, Glass K, Skrutkowska M. Monitoring clinical research: an obligation unfulfilled. Can Med Assoc J 1995; 152: 1973-1980.

Bridging section

In the preceding chapter, my colleagues and I studied a further step in the genesis and dissemination of medical knowledge, the reporting of eligibility criteria in clinical trial communications. Subject selection procedures must be accurately reported in order to allow for replication (or refutation) of research results, and, perhaps even more importantly, to allow clinicians in practice to determine to whom study results apply. We, therefore, sought to answer two questions: Are eligibility criteria accurately reported in communications regarding the results of clinical trials? If reporting is not complete, what is the nature of information loss (i.e., what types of criteria are not reported)?

We examined the reporting of eight clinical trials deemed important enough by the U.S. NIH to issue a Clinical Alert. For each clinical trial, we counted and classified (according to the schema developed in chapter 2) eligibility criteria in the protocol, methods paper (if applicable), journal article, and Clinical Alert. We observed a step-wise decline in the amount of information present at each stage of reporting: on average, methods papers contained 82% of the criteria listed in the protocol; journal articles, 63%; and Clinical Alerts, 19%. Reporting of 'definition of disease' criteria tended to be complete, whereas 'precision' and 'safety' criteria — categories of criteria relevant to clinicians — were reported incompletely.

Clearly, journal editors and persons responsible for issuing Clinical Alerts

need to ensure that clinicians receive all necessary information regarding clinical trial methods, including a complete list of eligibility criteria. If eligibility criteria are important they must be reported; if individual criteria are of questionable significance, they ought to be cut when the trial is designed, not when it is reported. Removing criteria of dubious necessity at the early stage will speed trial enrollment and help the trial mirror the range of patient in clinical practice (i.e., improve the study's generalizability).

In the preceding three chapters, we have examined clinical trial selection procedures at three different points: study protocol, interpretation by investigators, and reporting in trial communications. In the next chapter I stand back from these fine-grained empirical examinations and engage in a broad philosophical analysis of justice and the selection of subjects for clinical research. In particular, I ask: What are the implications of Walzer's political philosophy for the just selection of subjects for research participation?

Chapter 5:

Selecting subjects for participation in clinical research:

one sphere of justice

Abstract

NIH Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research (1994) prohibit the unjust exclusion of such groups from clinical research. But the requirement begs the question: What is a just selection procedure for participation in clinical research? In this paper, I outline an approach to this question based on the view of the complex egalitarian society articulated by Michael Walzer in *Spheres of Justice.*

I argue that a just selection procedure for participation in research is governed by the rule: eligibility criteria must select subjects solely in accord with the exigencies of medical care and science. Since the knowledge gained from research furthers the social good of medical care, such knowledge must be applicable to all members of the community. If the community is to be served, eligibility criteria ought to be minimized and necessary exclusions ought to be explicitly justified.

Ultimately, current policy is too narrowly construed. Once injustice is seen as (what Walzer calls) domination, it is clear that many groups are potentially affected: not only women and minorities, but also the elderly, the impoverished, the undereducated and the politically disempowered. All of these groups must be our concern. So too, the approach taken by current NIH regulations are unduly limited in their focus on publically-funded research. If the provision of medical

care is recognized as a social good, then all research in support of this good, including that which is privately funded, is subject to the justice considerations as outlined in this paper.

(251 words)

Introduction

Clinical research is the keystone in the provision of medical care. Clinical trials -- carefully designed studies of new treatments in human subjects -- are essential to ensure that therapeutic interventions are both safe and effective. But, important as it is, such knowledge cannot be gained legitimately at any cost. Research involving human subjects is governed by ethical principles laid out in the *Belmont Report*: respect for persons (the autonomy of individuals must be respected; those who are unable to make autonomous choices must be protected), beneficence (do no harm; maximize potential benefits and minimize potential harms), and justice (distribute the burdens and benefits of research fairly).[1]

One set of ethical issues involves the selection of subjects for participation in medical research. Selection procedures for research are determined by criteria for trial eligibility -- a check list of requirements set out in the study protocol that each prospective research subject must satisfy. For at least the last three decades, justice-related issues have surrounded subject selection for medical research. Rooted in conventional views of distributive justice, concerns, until recently, have focussed on the unequal allocation of the burdens and benefits of research participation. The exact nature of these concerns, however, has changed over time.

In the 1970s, the principle that the burdens of research participation should equitably distributed was dominant.[2] If "vulnerable" subjects ought to be protected

from research, the safest strategy was to exclude them from study participation. A number of factors contributed to this concern. Adverse events in trials to test new drugs, including the thalidomide disaster in the late 1950s (and early 1960s), engendered the perception that taking part in research was a risky endeavour. Furthermore, between 1965 and 1972, a number of research scandals came to the public's attention, each of which involved illicit experimentation on vulnerable groups.[3]

In the 1980s, a new primary concern emerged: the benefits of research participation ought to be equitably distributed. As a result of this shift, the presumption in the selection of subjects for research swung from one of exclusion to inclusion. What set this change into motion? *Contra* the perception that research participation involved substantial risk, studies published in the late 1970s and early 1980s revealed that earlier risk estimates were substantially exaggerated.[4] Undoubtedly, though, the main catalyst for the change was the advent of HIV/AIDS in the early 1980s. Given the absence of effective treatment for the disease in the first years of the epidemic, experimental treatments were seen as potentially life saving. Robert J. Levine observed that

> what was once seen as threatening -- a burden from which people would wish to be protected -- is now seen as a benefit. People are clamouring for access to clinical trials and to experimental drugs.

> People are demanding that they, and others who are like them, are owed such as a matter of justice.[5]

In the 1990s, while issues of the equitable distribution of both the burdens and benefits of research remain important, a new object of ethical focus has emerged: knowledge gained from medical research must be fairly distributed, i.e., the results of research studies ought to be widely applicable. As Levine has pointed out, this new concern should not be confused with the above (1980s) issue: equitable distribution of benefit demands that "*individuals* ought not to be excluded from research and its associated *direct benefits* [emphasis added]."[6] The new ethical dictum requires that "*[classes of] persons* ought not to be deprived of the benefits of research, *generalizable knowledge*, and the development of new therapies [emphasis added]."[7]

Originally the concern evolved out of the fact that classes of individuals, including women, children and intravenous drug users, were barred access by study eligibility criteria to clinical trials of treatments for HIV/AIDS. The exclusion of these groups, according to Macklin and Friedland, led to

> a lack of information on the efficacy of AZT on a wider group of AIDS patients, stemming from the fact that the demography of patients studied did not replicate the entire population of individuals with AIDS.[8]

The issue expanded rapidly beyond the confines of HIV/AIDS research. In 1992, Rebecca Dresser made the controversial claim that "the failure to include women [and minorities] in research populations is ubiquitous."[9] As the issue has picked up speed in the political arena, a number of changes in federal research policy have been made. In 1993, the U.S. Food and Drug Administration (FDA) removed barriers to the enrollment of women "of reproductive potential" in early studies evaluating new drugs.[10] In 1994, the U.S. National Institutes of Health (NIH) released the *NIH Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research.*[11] In part, these guidelines require the representative inclusion of women and minorities in all NIH-funded research.

U.S. Institutional Review Boards (IRBs), local committees that review human subjects research for ethical acceptability, have been given part of the responsibility to enforce these recent policy changes.[12] When reviewing a study for ethical acceptability, IRBs must ensure that adequate representation of women and minorities is likely. In part, this involves scrutinizing eligibility criteria to ensure that groups have not been excluded inappropriately. Also, IRBs may: review procedures for study enrollment, require the development of programs to ensure continued participation of women and minorities in studies, and review demographic information on study enrollment on an annual basis.

In order to implement these rules optimally, however, a comprehensive

theoretical construct is required. Several feminist theorists haven broken sharply with conventional approaches to justice-related issues and provide useful insights into justice and subject selection for research. Rather than focussing on the equal distribution of goods (e.g., research benefits and risks -- a "negative" good), these theorists take the *elimination of oppression* to be the key issue in discussions of justice. Sherwin points out that feminist discussions of justice in research share the concern that the exclusion of certain groups from research is a manifestation of societal oppression.[13] DeBruin argues that oppression of women manifests itself in at least two general ways: the male experience is taken to be the norm (and thus women, and female physiology, are seen as abnormal), and women as a group, and health issues particular to women, have been subordinated to the dominant group (men and male health concerns).[14] A just approach to the allocation of knowledge derived from research must, according to this view, aim towards eliminating oppression. With regard to subject selection, this entails enrolling sufficient numbers of women in studies so that the results can be meaningfully applied to women. But Sherwin and DeBruin believe that more than this is required: preferential attention to women's health problems must be given in order to make up for the "history of gender-based oppression" in health research.

The introduction of the concept of 'oppression' into discussions of justice is, I think, a productive step, but it seems to take us only part way to a

comprehensive approach to the just selection of subjects for research participation. A number of key questions have not, to date, been addressed. What counts as an oppressed group? Oppression is a term that is, at best, vaguely defined. Children were excluded from early studies of the treatment of HIV/AIDS, but are children an oppressed group in society? (Perhaps so, perhaps not.) How to account for the problematic exclusion of non-oppressed groups? An oppression-based theory seems unable to address the exclusion of groups that are not oppressed. What if a highly-idiosyncratic investigator decided to exclude all members of the Conservative Party from a research study? Clearly, this seems unjust, but on what basis is it so? Furthermore, if members of oppressed groups may be excluded legitimately from a study, what counts as an adequate justification? Finally, and most important, "What does a just eligibility criterion look like?" In order to evaluate comprehensively criteria for trial eligibility, we need a "big-picture" account of what a *just* selection procedure involves.

Without rejecting a feminist approach to the problem, in this paper I want to provide an alternative (and possibly complementary) approach using Michael Walzer's theory of justice as articulated in *Spheres of Justice*.[15] Walzer's view of injustice as domination may allow for a more comprehensive account of justice in subject selection than is currently available. Since I focus on justice and the selection of subjects for medical research, I will not discuss the impact of Walzer's

views on larger health policy issues, including the prioritization of research initiatives.

Spheres of Justice

Michael Walzer's approach to questions of distributive justice is, I think it fair to say, a radical departure from conventional theories of justice. Equality, says Walzer, is not about ensuring that everyone has an equal number of things (i.e., simple equality); rather, its aim is to free society from domination. At root, the driving motivation for equality is the experience of subordination:

> What is at stake is the ability of a group of people to dominate their fellows. It's not the fact that there are rich and poor that generates egalitarian politics but the fact that the rich 'grind the faces of the poor', impose their poverty upon them, command their deferential behavior.[16]

For Walzer, a just society is not one in which simple equality exists, but rather one in which the means of domination has been eliminated. Although the specifics may vary from one time and place to another, domination is always mediated by some social good, e.g., money, political office or birth and blood.

Walzer sets himself the task of describing a society that has been freed of domination, a complex egalitarian society. The construction of such a society

cannot be detached from historical and cultural particularities:

> [T]he question most likely to arise in the minds of the members of a political community is not, What would rational individuals choose under universalizing conditions of such-and-such a sort? But rather, What would individuals like us choose, who are situated as we are, who share a culture and are determined to go on sharing it? And this is a question that is readily transformed into, What choices have we already made in the course of our common life? What understandings do we (really) share?[17]

In order to understand how particular goods ought to be distributed in a particular society, we must understand what those goods mean to members of that society. A bit of food may represent sustenance or a religious offering; and what is to be done with it depends on which of these meanings it is given. Shorter: goods "come into people's minds before they come into their hands."[18]

How shall we understand goods? Walzer offers us six propositions regarding goods.[19] First, goods are irreducibly social in nature; they cannot be held idiosyncratically. Second, the distribution of goods cannot be understood in the abstract; the actors in distributive transactions behave the way they do because of how they conceive certain goods. Third, there is no single set of primary goods across all moral and material worlds; the range of even necessities and their rank

orderings differ too widely to define usefully a universal basis for distribution. Fourth, just distributions are determined by the meaning of goods in a particular community. Fifth, the meaning of goods within particular communities is historical in character and, thus, is subject to change over time. Sixth, and most importantly, when meanings are distinct, distributions are autonomous; the distribution of a particular social good operates within "a distributive sphere within which only certain criteria and arrangements are appropriate."[20]

Thus, for Walzer, the just distribution of social goods cannot be understood by a rule universal in application, governing all transactions. Each of the social goods -- community membership, security and welfare, money and commodities, office, hard work (a "negative" good), free time, education, kinship and love, divine grace, recognition, and political power -- operates within its own autonomous sphere. Each sphere is governed by separate rules of just distribution. Given this understanding, Walzer defines domination as a transgression of the boundaries between (or among) spheres of justice. Domination exists when one good is converted into another when there exists no intrinsic connection between the two. Complex equality, then, establishes social relationships within a community such that "no citizen's standing in one sphere or with regard to one social good can be undercut by his standing in some other sphere, with regard to some other social good."[21] More formally: "*No social good x should be distributed*

to men and women who possess some other good y merely because they possess y and without regard to the meaning of x."[22]

Walzer distinguishes dominance from monopoly. Dominance, as we have seen, exists when one good commands control of other goods or a wide range of goods. Monopoly exists when a single individual, or group of individuals, owns or controls that good. Thus,

> [d]ominance describes a way of using social goods that isn't limited by their intrinsic meanings or that shapes those meanings in its own image. Monopoly describes a way of owning or controlling social goods in order to exploit their dominance.[23]

Conventional approaches to distributive justice have focussed on eliminating monopoly but not dominance: they ask, How can this good be distributed more fairly (usually meaning more equally)? Money is a dominant good within our society (it can buy education, health care, political office, and, yes, even respect -- none of which, according to Walzer, are legitimate transactions), and one that is monopolized by a relatively small group of individuals. Merely redistributing wealth would be a false, or at best a temporary, solution. Some people would save their money, others would spend it, still others would by their own entrepreneurship accumulate wealth; and so, the distribution would very soon be unequal again. If government intervened repeatedly to ensure an equal distribution,

money would no longer be a dominant good, political power would be.

As opposed to conventional approaches to distributive justice, Walzer's solution is to address dominance directly. Walzer's approach is to narrow the range of legitimate operation of social goods such that they are only convertible within their own individual spheres of operation. But this cannot be derived from the abstract, one must examine the meanings of various social goods and their just distribution for a particular community.

> It will yield not an ideal map or a master plan but, rather, a map and a plan appropriate to the people for whom it is drawn, whose common life it reflects. The goal, of course, is a reflection of a special kind, which picks up those deeper understandings of social goods which are not necessarily mirrored in the everyday practice of dominance and monopoly.[24]

The complex egalitarian community will bar so-called "boundary crossings" (converting goods between spheres), but will not prevent unequal distributions of individual goods. The point is, though, that in such a society:

> Though there will be many small inequalities, inequality will not be multiplied through the conversion process. Nor will it be summed across different goods, because the autonomy of distributions will tend to produce a variety of local monopolies, held by different

groups of men and women.[25]

Medical care

How does Walzer approach justice and medical care? In Walzer's typology, medical care is classified in the sphere of security and welfare, which also includes provision for the poor, public works projects, defence, and securing a supply of food and trade generally. The mutual provision of needs is undoubtedly the most basic requirement of the community. Members of a community have obligations to participate in the provision of such needs to one another, and, as such, they have obligations to fellow members which go above and beyond obligations to strangers. Walzer explains that

> [C]ommunal provision is important because it teaches us the value of membership. If we did not provide for one another, if we recognized no distinction between members and strangers, we would have no reason to form and maintain political communities...Political community for the sake of provision, provision for the sake of community: the process works both ways, and that is perhaps its crucial feature.[26]

But which needs must the community provide for? Separate communities at different times have answered this question in disparate ways: for some common

defence is a priority and caring for the poor and the sick is not, for others, the reverse is true. Priorities can change over time as well. In Europe in the middle ages, salvation was an all-important public good, and one that was communally provided for: churches were built in every parish, regular services were held, communion was compulsory and so on. The desire for eternal life, in Europe and North America at least, has come to be replaced with the more profane need for long (earthly) life -- longevity.

> Among modern citizens, longevity is a socially recognized need; and increasingly every effort is made to see that it is widely and equally distributed, that every citizen has an equal chance at a long and healthy life: hence doctors and hospitals in every district, regular check-ups, health education for the young, compulsory vaccination, and so on.[27]

The point is that the priority of needs is determined by the community itself and is based on the community's conception of the common life. The decision by the community as to which goods are required by the common life is the essence of the social contract. Once these needs are defined, the community must ensure that the needs are attended to; "that the goods that are distributed must be distributed in proportion to need; and that the distribution must recognize and uphold the underlying equality of membership."[28]

In contemporary Canada and the United States, medical care is seen as an important social good. But providing medical care and conducting medical research is expensive and, therefore, communal effort is required. The community has stepped in to pay for basic medical care (*via* Medicare and Medicaid in the United States; *via* provincial health care plans in Canada). But

> once communal provision begins, it is subject to further moral constraints: it must provide what is 'wanted' equally to all members of the community; and it must do so in ways that respect their membership.[29]

Given that medical care is an important social good, providing such care on the basis of wealth (the sphere of money and commodities) involves "boundary crossing" and is therefore unjust. The distributive principle of health care is, *Medical care must be proportionate to illness.*

But an interesting problem arises from the fact that the provision of medical care may lead to a number of goods, and these goods may originate from a number of "competing spheres." Medical care will, as described above, provide aid to the ill, but it will also provide other benefits. Health care workers will derive incomes from their work; skillful practitioners will derive prestige from their work and earn the respect of others; and, in turn, this prestige may translate into other goods such as political office. How are we to determine the sphere which properly determines

the just distribution of health care?

This difficult question is not one that Walzer addresses directly. But, I think it safe to say, that if we acknowledge that the distribution of a good is determined by its meaning to a community and if that meaning can change over time, then a given good may fall into different spheres in different communities, or in the same community at different times. For example, when medical care, as described above, was not a priority, and was, therefore, not communally provided for, its just distribution was governed by the rules set for commodities rather than elements of the sphere of security and welfare.

When several spheres seem to compete for a single good, we must scrutinize the meaning of *that* good for *this* community to determine its primary meaning and thus its primary distributive rules. Within our society, health care is communally provided for because health and longevity itself -- not the income, prestige, or career advancement of, for example, physicians -- is a communal priority. We can imagine communities in which these other goods might be key, but it seems unlikely that health care would be provided communally in them.

Medical research

Walzer say relatively little about medical research specifically: he identifies it as a part of the sphere of security and welfare and he indicates that it is a part of

the common effort required to provide medical care.[30] As we have indicated, medical research is necessary to progress in the fight against disease. Basic research helps us understand basic physiological processes that are disturbed by disease states, dissect the workings of the agents of disease, and develop and test new treatments in (non-human) model systems. Clinical research, resting on the foundation of basic research, tests new treatments in human subjects in order to ensure that novel interventions are both safe and effective. Without medical research we could neither be sure that current treatments actually work nor develop effective treatments for (currently) untreatable disease. Medical research is perhaps viewed best as a necessary means to the provision of medical care, rather than a good in itself.

But it need not be so and, indeed, it has not always been so. Our society's changing perceptions of cancer treatment and research provides an instructive example. Patterson, in his scholarly social history of cancer research and treatment, *The Dread Disease: Cancer and Modern American Culture*, points out that cancer, until the early part of this century, was largely viewed as an untreatable disease.[31] Skepticism regarding the worth of seeking treatment from physicians in general was wide spread at the time; physicians, in North America at least, were often poorly trained and had few effective treatments to offer patients. Medical science too seemed to have little to contribute:

> With medical science in such an uncertain and defensive state, the topic of cancer was riddled with controversy and bewilderment in the 1880s and 1890s. At the root of the bewilderment was one intractable reality — scientists did not know how or why cancer cells broke loose on their destructive paths.[32]

Perhaps not surprisingly, biological research was not publicly funded at the time. Skepticism about medical treatment was such that even if cancer research had been publicly funded, one can well imagine the end of such funding being to further knowledge as an end separate from improving treatment for the disease. In a medically fatalistic society, as was 19th century America, research into disease could remain unconnected from medical treatment -- if the outcome of a disease is fated, science can only describe, not alter the inevitable. In such a society, research would be viewed in a sphere of operation -- namely, that of knowledge and education -- apart from that of medical care.

After the turn of the century, a number of striking medical advances, including Erhlich's discovery of Salvarsan, the first effective treatment for syphilis, began to change the public's perception of medical treatment. The medical profession too changed: important reforms to medical education at the turn of the century helped standardize the training that physicians received. As these changes took root, Americans came to accept the "medicalization" of

diseases like cancer, and came to invest in science their hopes for progress in the fight against disease:

> The growing prestige of physicians reflected significant social differences in the way that Americans perceived organized medicine and diseases, cancer included. These differences may indeed have grown as a result of the new claims for scientific medicine...[M]any Americans of means came to believe the claim of the experts. They could afford to see physicians and to pay for hospital rooms, and they agreed to be treated there. They accepted the modern medical culture.[33]

As cancer became medicalized, the connection between improvements in medical treatment and medical science became solidified: cancer was seen as "both a looming threat to civilization and a disease that brilliant scientists were beginning to conquer."[34] But if it was to be conquered, concerted government action was required. In July 1937, legislation was passed creating the U.S. National Cancer Institute. The new Institute would

> conduct researches, investigations, experiments, and studies relating to the cause, diagnosis, and treatment of cancer...with a view to the development and prompt widespread use of the most effective methods of prevention, diagnosis, and treatment of cancer...[35]

Cancer research was publicly supported to the end of *improving treatment for the disease* and subsequent funding measures, including President Nixon's 1971 declaration of a war against cancer, have been solely justified and accepted on this basis.[36] Thus, in our society, publicly-funded cancer research -- and, I believe, a similar story may be told for other categories of medical research -- is subservient to the goal of the communal provision of effective medical care.

If medical research is necessary to the communal provision of effective medical care, what steps should be taken by the community to ensure an adequate supply of research subjects? Enrolling sufficient numbers of human subjects in a timely fashion into clinical research has been identified as a major problem in, for example, cancer research.[37] Part of the problem is that a relatively small proportion of cancer patients is actually enrolled in research studies. Indeed, in the United States only 1.6% of cancer patients are treated in trials funded by the National Cancer Institute.[38] Enrolling adequate numbers of human subjects in trials is essential to answering study questions efficiently, and, therefore, it is key to progress in the fight against disease. How ought a just society ensure that adequate numbers of research subjects are available for study? Three approaches suggest themselves: conscription, payment, and volunteering.[39]

Conscription (or perhaps a "research-participation tax") seems like an obvious solution to the problem. In order to further the communal provision of

medical care, each member of the community might be required to participate in a certain number of hours of research per year. Since some research is more demanding (more tests or invasive procedures, greater risk), the time commitment required might be weighted accordingly. But conscription goes too far: paying taxes to support medical research is one thing, being required to undergo biopsies or be exposed to potentially hazardous drugs, quite another. In short, the proposal goes against our common understanding of the importance of bodily integrity.

Paying research subjects salaries for their participation represents another possibility. Various public projects, including the building of roads, dams and public buildings, require labor from the community to ensure their completion. While in the past the state may have employed conscription or forced labor to complete such projects, in contemporary society labor is secured by paying people for their work. Why not pay people to participate in research studies? Given sufficient remuneration, even the riskiest study is likely to attract sufficient participants to ensure speedy completion.

There are at least three problems with this suggestion. First, the lure of rich reward may cause potential subjects to withhold critical medical information that would make them ineligible for study participation. In several cases, deaths of "healthy" volunteers in phase I studies have resulted from the fact that subjects have withheld such information, seemingly because they needed (or wanted) the

money offered for participation.[40] Second, the poor (and unskilled) may be lured disproportionately by the money offered in research. If the provision of medical care on the basis of ability to pay is "boundary crossing" (i.e., unjust), then so too is the disproportionate enrollment of the poor in research studies. Finally, payment, particularly if the amount is substantial, conflicts with a basic tenet of ethical experimentation: the *voluntary consent* of the research subject is required. According to the *Belmont Report*,

> An agreement to participate in research constitutes a valid consent only if voluntarily given. This element of informed consent requires conditions free of coercion and undue influence. Coercion occurs when an overt threat of harm is intentionally presented by one person to another in order to obtain compliance. *Undue influence, by contrast, occurs through an offer of an excessive, unwarranted, inappropriate or improper reward or other overture in order to obtain compliance* [emphasis added].[41]

Ultimately, clinical research must depend upon the altruism of individuals who volunteer for research participation. Steps can and indeed ought to be taken to encourage research participation: expenses, such as transportation, incurred by research volunteers should be reimbursed, and the community should acknowledge the important contribution made by volunteers. Altruism is something that the

community ought to encourage. As Walzer points out, the act of giving is a good in itself: it fosters a closer connection between the individual and her community.[42]

> One might think of the gift relationship as a kind of politics: like the vote, the petition, and the demonstration, the gift is a way of giving concrete meaning to the union of citizens. And as welfare generally aims at overcoming the dominance of money in the sphere of need, so the active participation of citizens in the business of welfare (and security, too) aims at making sure that the dominance of money is not simply replaced by the dominance of political power.[43]

The distributive logic of medical research

We have argued that medical research is necessary to the provision of high-quality medical care and that it is funded to this end in our society. If the distribution of knowledge generated from medical research is to "recognize and uphold the underlying equality of membership",[44] the results of medical research must be applicable to the breadth of community members afflicted with a particular illness. If the results of phase III clinical trials (final-stage studies aimed at changing medical practice) are to be generalizable, i.e., applicable to the community at large, then the population studied within that trial ought to be representative of the community (of affected individuals) at large. Thus, an

inclusive stance to subject selection must be adopted and the exclusion of particular groups or types of patients from studies must be carefully justified.

But some eligibility criteria are required. If a phase III clinical trial has no criteria defining patient eligibility, meaningful results can not be generated: one cannot know to whom the results of the trial apply. Just eligibility criteria will be those that are essential to further the end of the trial, namely, to produce results that advance medical care in the community at large. The distributive logic of the selection of subjects for trials is this: *eligibility criteria must select subjects solely in accord with the exigencies of medical care and science.* If an eligibility criterion selects (or excludes) subjects because of their standing in another sphere, without reference to the requirements of medical care or science, the criterion involves "boundary crossing" and is unjust. Clearly, if the justification for individual eligibility criteria are to be open to scrutiny, they must be made explicit by trial designers in study protocols.

What are some examples of just eligibility criteria? Our recent study of eligibility criteria in two sets of important clinical trials is a good source of examples.[45] In the study we developed a schema for the classification of eligibility criteria that divided criteria into five categories:

- *definition of disease*: eligibility criteria that define the medical condition of interest and represent factors that would be taken into account in clinical practice;

● *precision*: eligibility criteria concerned with the scientific validity of the study. These criteria attempt to diminish variability in the study by either making the patient population more homogeneous or reducing measurement error. Precision criteria involve factors that would not ordinarily be taken into account in clinical practice;

● *safety*: eligibility criteria that exclude persons thought to be unduly vulnerable to treatment in general or one of the study treatments in particular;

● *ethical and legal*: eligibility criteria that are required in order to ensure conformity with Department of Health and Human Services (DHHS) regulations governing the conduct of human experimentation;

● *administrative*: eligibility criteria which attempt to ensure the smooth functioning of the study. Administrative criteria such as measures aimed at ensuring compliance with treatment and follow-up fall into this category.

Let us consider some examples of eligibility criteria from the clinical trials of the National Surgical Adjuvant Breast and Bowel Project (NSABP) examining the treatment of stage II, node positive (locally advanced) breast cancer.

In the NSABP breast cancer trials, eligibility criteria fell into the five categories as follows: definition of disease, 41%; precision, 36%; safety, 11%; ethical and legal, 9%; and administrative criteria, 3%. In so far as definition of disease eligibility criteria formalize clinical decision making, they are both

necessary and justified restrictions of the study population. For example, criteria that require that "the tumor is confined to the breast or axilla", the patient have "one or more positive axillary nodes proven histologically" and not have "inflammatory carcinoma" reflect the exigencies of medical care. Other categories of criteria clearly reflect the exigencies of science. Administrative criteria, for example, "patient is accessible (geographically) for follow-up," mirror the fact that the study must accrue data on subjects in order to reach a successful conclusion. Still, these criteria must be justified: it is not enough to suppose that one group of patients is non-compliant (this view may be the result of widely-held, but false, belief about a group), evidence must be marshalled to support such a claim. Criteria that fulfill ethical and legal requirements are also required by science; they define the proper conduct of science. For example, each of the breast cancer studies required that the "patient consents to be included in the study."

Other categories of criteria do not hold up so well to the presumption of inclusiveness (and generalizability). In our previous paper we operationalized the inclusiveness requirement as follows, "[e]ligibility criteria in phase III clinical trials should restrict the eligible patient population as little as possible, consonant with the demands of scientific validity." The NSABP studies contained numerous precision criteria that excluded patients with "previous malignancy," with "concomitant disease," with serious "non-malignant systemic disease," who

received prior "irradiation...[or] chemotherapy," or who received "prior hormonal therapy." While each of the criteria may serve the narrow "scientific" goal of making the study population more homogeneous (i.e., less variable), it is at the expense of the applicability of study results to the community at large. Similarly, safety criteria in the series of trials seemed to exclude some groups of patients who would indeed require treatment in clinical practice with agents similar to those in the clinical trials. Studies routinely excluded patients with abnormal renal, hematological or hepatic indices. Patients with mild or moderate disturbances in these indices certainly require treatment. In our prior paper we point out

> Exclusion of these groups leaves the clinician with no information on the risks and benefits of investigational treatments in such groups. The clinician may well wonder: Is a woman with breast cancer who has received prior cancer treatment more susceptible to harmful effects from treatment? Do the benefits of the proposed treatment outweigh these risks? Unfortunately, no information can be forthcoming from trials that exclude such groups.[46]

Issues within the sphere in which medical research operates must pass the test of the distributive logic particular to research: *eligibility criteria must select subjects solely in accord with the exigencies of medical care and science*. But medical research does not operate within a vacuum: it serves the social good of

medical care and thus the community as a whole. Thus, justice issues within the sphere predominantly involve the issue of generalizability: to whom do the results of research apply?

A natural tension exists between research studies that are widely generalizable and those that are scientifically fastidious. A research protocol with a highly-restrictive set of eligibility criteria may, indeed, lead to a study population that is relatively homogenous. As a result, it is at least conceivable that a more precise estimate of the outcome measure will result from a fastidious study than one that is widely inclusive. A fastidious approach is likely to appeal to clinician-scientists who embrace the *ethos* (and, indeed, the aesthetic) of the controlled laboratory experiment: control for all factors except the independent variable in question.[47] But such narrowly-focussed studies produce results that are only applicable to a narrow segment of the patient population (i.e., those few who would have been eligible for study participation).

A Walzerian analysis forces us to acknowledge that the prime purpose, the *telos* if you will, of clinical research is to further the medical care of all afflicted community members. Clinical researchers who wish to unduly restrict study entry may be motivated by a variety of factors. As we have said, they may be moved by the aesthetic of the highly-restrictive trial. Indirectly, a career based on such trials, perceived by many to be elegant, may bring the investigator prestige and

recognition. A career built on restrictive trials may also lead to academic advancement and greater political influence among funding agencies and other clinician-scientists with similar trial aesthetics. Finally, the investigator may believe that narrowly-conceived trials produce "higher-quality knowledge," even though the knowledge gained is less clinically applicable. But each of these considerations originates from spheres separate from that of security and welfare (medical care): concerns over aesthetics and recognition originate from the sphere of recognition; career advancement and influence, from the sphere of political office; and knowledge *simpliciter* (i.e., knowledge apart from clinically-relevant knowledge) as an end, from the sphere of education. Allowing these concerns to reign in clinical research is domination, and hence the factors must be rejected as primary motivators for selecting subjects for research. Eligibility criteria that heed the distributive rules of the sphere of security and welfare will minimize restrictions to the study populations, and will justify necessary restrictions carefully and in a manner open to public scrutiny (e.g., review by IRBs).

Unjust inclusion as domination

Until this point we have focussed on defining the rule for the just distribution of knowledge resulting from clinical research within the sphere of welfare and security. Another category of justice relevant to human

experimentation exists: justice as the absence of domination. Walzer, as we have seen, characterizes domination as boundary crossing. In the context of medical research, domination occurs when an eligibility criterion selects (or excludes) subjects because of their standing in another sphere, without reference to the requirements of medical care or science. To take the example from the introduction, a study that bars Conservatives from participation excludes subjects on the basis of their standing in the sphere of political power (and without reference to the exigencies of medical care or science). Such a requirement involves domination and is therefore unjust. So too, *mutatis mutandis*, medical research is unjust if it (without regard to medical or scientific requirements) selects subjects of the basis of wealth (sphere of money and commodities), education (sphere of educations), political empowerment (sphere of political power), gender (sphere of kinship and love), citizenship (sphere of membership), employment (sphere of hard work) or religious belief (sphere of divine grace).

As indicated in the introduction, a number of prominent research scandals in the last fifty years have involved the exploitation of vulnerable groups in society. The ethical violations in many of these cases were so all-encompassing as to breech all the basic principles of research: respect for persons, beneficence and justice.[48] In these studies, meaningful informed consent was often not obtained, and the research presented an inordinate amount of risk in relation to benefit (if

any). In so far as these experiments were unjust, though, they can be usefully characterized as instances of domination.

Perhaps the best known example of unethical research is the experimentation carried out by the Nazis on captive populations during the Second World War.[49] Many of the Nazi experiments were motivated by the exigencies of war: they examined the treatment of hypothermia, battle wounds and infectious disease. No informed consent was obtained from research participants (or consent was obtained on false grounds), the experiments often resulted in the death of the research subject (approximately 25% of the subjects in the hypothermia experiments at Dachau died as a direct result of the research), and even basic principles of scientific design and validity were violated.[50] The subjects for these experiments were largely political prisoners deemed expendable by the state. For example, the hypothermia experiments at Dachau "recruited" Polish and Russian political prisoners for the various studies that were carried out. The selection of subjects is, in this case, a clear example of domination: persons were included in research by virtue of their status as political prisoner (or politically disempowered persons) -- that is, by virtue of their standing in the sphere of political power -- and without reference to the exigencies of medicine or science.

The Tuskegee syphilis study is another well known example of unethical research. Perhaps the longest running study funded by the U.S. Public Health

Service (it was active between 1932 and 1972), the Tuskegee syphilis study enrolled roughly 400 Afro-American males from rural Alabama in to order study the effects of untreated syphilis.[51] While therapy for syphilis was toxic and relatively ineffective when the study began, study participants were not given access (and in some cases were denied access) to penicillin -- a non-toxic and highly effective treatment for the disease -- when it became available in the early 1950s. As a result, it is estimated that 20% of the participants died from the complications of syphilis. Also, proper informed consent was not obtained in many cases: subjects were not informed that they had syphilis, they were told they had "bad blood"; procedures done purely for research, such as spinal taps, were described to subjects as "treatments". The study was unjust because it exclusively enrolled subjects who were poor, uneducated and Afro-American. Subjects were selected (at least in part) by virtue of their standing in the spheres of money and commodities, education, and political power. It is, therefore, a clear example of injustice as domination.

Unjust exclusion as domination

But domination can cut both ways: subjects can be unjustly included in research, or they can be unjustly excluded. Recent interest has focussed on the latter rather than the former. When domination takes the form of unjust inclusion

in studies, subjects are exposed to the risks associated with the research; when domination takes the form of unjust exclusion, other "harms" may be incurred. Insufficient information may exist to ensure that such groups within society receive effective medical care: members of excluded groups may be exposed to ineffective treatments, unexpected side-effects may occur, or, more generally, a lack of information may lead to delays in the diagnosis and treatment of disease.[52] Two groups in society, women and the elderly, allow us to examine both of these components: Have they been excluded from research? Has the lack of medical research affected the medical care provided?

Women. The claim has been made that women are excluded from a wide range of clinical research. Women have been excluded from research studies for a variety of reasons: male physiology is taken to be the norm and menstrual cycles are a "complicating factor", women may be harder to enroll and retain in research studies, and, if a women becomes pregnant, harm to the fetus may ensue.[53]

Evidence for the *ubiquitous* exclusion of women from research is, however, lacking. According to 1977 U.S. Food and Drug Administration Policy (rescinded in 1993), all women of reproductive potential must be excluded from early stage testing of new drugs.[54] This exclusion does not seem to have carried over into later stage clinical research. In 1992, the U.S. General Accounting Office (GAO) reviewed the phase II and III clinical research supporting New Drug Approvals

issued between January 1988 and June 1991.[55] Defining inadequate gender representation as <40% of the study population, all classes of drugs, except new cardiovascular drugs, were supported by studies with adequate gender representation. Even women between 15 and 49 years of age (women of so-called "reproductive potential) were not generally underrepresented. Bird reviewed research articles published in *JAMA* in 1990 and 1992.[56] Defining gender underrepresentation as blanket exclusion or less than one-third of the study sample (for disease that affect both genders), women were underrepresented in 2.7 times as many studies as men. Once again, though, the strongest evidence came from cardiovascular research studies (only research in cardiovascular disease and substance abuse underrepresented women in the majority of studies).

In other areas of research it seems quite clear that women have *not* been excluded from research participation. Ungerleider and colleague reviewed 1989 accrual data to cancer studies funded by the Clinical Trials Cooperative Group Program of the U.S. National Cancer Institute.[57] Of the 18411 subjects enrolled in cancer clinical trials, 57% were women. (When pediatric cooperative groups were excluded, 60% were women; when cooperative groups that predominantly study cancers specific to women were excluded, 45% were women). In an extensive review of the literature, the Committee on the Ethical and Legal Issues Relating to the Inclusion of Women in Clinical Studies identify the "lack of reliable,

comprehensive information on the actual participation of women...in clinical studies."[58] Indeed, the Committee was only able to find evidence that women were excluded from studies of heart disease and (perhaps) HIV/AIDS.

If women have been largely excluded from research on heart disease, has this had a negative impact on the medical care that women with heart disease receive? It seems the answer may be yes. Ayanian and colleague studied the medical records of men and women hospitalized with coronary artery disease in 1987 in Massachusetts and Maryland.[59] Despite the fact that gender is not (in itself) a predictive factor of heart disease, women underwent fewer major diagnostic and therapeutic procedures than men. Steingart and colleagues looked at the care of 1842 men and 389 women after an acute myocardial infarction (and before enrollment in the Survival and Ventricular Enlargement Trial).[60] Despite greater reported functional disability from coronary artery disease in study women, they were roughly half as likely to have undergone cardiac catheterization or coronary bypass surgery. Finally, the American Medical Association's Council on Ethical and Judicial Affairs concluded that heart disease is more advanced in women by the time they receive surgery and women have a higher operative mortality rate for coronary artery bypass.[61]

It is clear then that, at least with regard to heart disease and early-stage clinical trials, women have been excluded from participation in medical research.

Why have they been excluded? Two reasons have been given: female reproductive physiology is a "complicating factor" in scientific studies, and concern for toxicity of experimental treatments to (actual or potential) fetuses.[62] While exclusions may be legitimate in particular cases (for example, when pregnant women are excluded from a study of a drug known to be teratogenic), the wide-spread exclusion of women from studies on these grounds is, as we shall see, domination.

Walzer recognizes that the family is an important unit within contemporary society. Particular rules of distributive justice apply in what Walzer interchangeably calls the sphere of 'kinship and love' and that of 'personal relations, domestic life, reproduction and child rearing.' Women have an important role to play within the family, one that involves, *inter alia*, carrying a child to term, giving birth to it, and playing an important role in caring for it after birth. While this reproductive role has an undoubtedly central place in the family, women have been illegitimately excluded from other social roles on the basis of this familial role, on the basis of their standing within the sphere of kinship and love. This exclusion, according to Walzer, is domination and, therefore, unjust:

> Alongside nepotism — an expression of kinship preferences where preference has no proper place — there has long existed something like its opposite: a kind of political and economic misogyny — an expression of kinship constraints where constraint has no proper

> place. Thus the denial to women of the right to vote, or to hold office, or to own property, or to sue in court, and so on. In each case, the reasons given when anyone bothers to give reasons, have to do with woman's place within the family. So kinship patterns are dominant outside their sphere. And liberation begins outside, with a succession of claims that this or that social good should be distributed for its own, not for familial, reasons.[63]

The boilerplate exclusion of women of reproductive age and women in general from many scientific studies is based on the reproductive role that women fulfil within the family. As we have said, studies have excluded women because of the fact that the reproductive physiology of women is viewed as a "complicating factor."

> Male norm refers to the tendency to conceive of men gender neutrally, as persons, rather than to conceive of men in terms of their sex. Thus, men's identity and experience becomes the characterization or standard of what it is to be a human being. An example is the argument that men make more appropriate research subjects for drug studies because, for example, women have menstrual cycles that produce deviations from the 'normal' pattern of drug disposition observable in males.[64]

Although men and women share many more physiological similarities than differences, differences nonetheless do exist. The point here, though, is not whether differences exist, but rather that one gender be taken as the norm and the other excluded on the basis of 'reproductive standing.' This characterization of women -- and, importantly, *not men* — in terms of their reproductive role and their exclusion on this basis from scientific studies is an instance of domination and is unjust.

Women, more specifically women of reproductive age, have also been excluded from studies on the grounds that experimental drugs may be harmful to (potential or actual) fetuses. Toxicity to the fetus is certainly a legitimate concern in some cases. When a given drug is known to be teratogenic, excluding pregnant women from participation in the study (including requiring a pregnancy test at the beginning of the study and the use of effective contraception throughout the study) seems like a reasonable -- and even necessary -- precaution. But, even in such cases of clear risk to the fetus, the exclusion of *all women of reproductive age* from a study seems, as above, to characterize women solely on the basis of their reproductive role. Patterson and Emanuel describe two studies at the Dana-Farber Cancer Institute both of which involved drugs with the potential to harm the developing fetus: 13-*cis* retinoic acid (for the prevention of secondary lung cancers) in one trial, and finasteride (for prostate cancer) in the other.[65] In the 13-

cis retinoic acid study, all women of reproductive age were excluded; in the finasteride study, men were merely required to use an effective contraceptive method. Moreno, quoted in the discussion of the two cases, concludes that

> it seems fairly obvious that if we are going to trust men to use contraception even when their semen is contaminated with finasteride and might pose a risk to a developing male, then we certainly should trust women to use contraception when taking retinoic acid. I can see no substantive difference between these two situations which should lead to a difference in policies regarding the eligibility of patients for the trials.[66]

In both trials it was reasonable to take steps to protect fetuses from harm. The exclusion of *all* women on reproductive age in the one trial, however, clearly characterizes and excludes women on the basis of their reproductive role, on the basis of their standing within the sphere of kinship and love, and is therefore an instance of domination. If the exclusion of all women of reproductive potential is unjust in the case of a clearly teratogenic drug then, *a fortiori*, it is unjust when evidence for teratogenicity is less clear or lacking.

As we have seen, the exclusion of women from some areas of medical research, particularly research into cardiovascular disease, has affected their medical care. Healy comments that

> Decades of sex-exclusive research have reinforced the myth that coronary artery disease is a uniquely male affliction and have generated data sets in which men are the normative standard. The extrapolation of these male-generated findings to women has led in some cases to biased standards of care and has prevented the full consideration of several important aspects of coronary disease in women.[67]

If women are to receive an allocation of medical care which upholds "the underlying equality of membership" in the community,[68] they must be included in medical research. Criteria that exclude women for "scientific reasons" are often baseless exclusions and instances of domination. Clinical trial designers must provide clear justifications for any exclusion and evidence to support such exclusions ought to be carefully examined.

Even when strong evidence exists that, for example, women may respond differently than men to a particular treatment (I take this to be the exception), the equality provision does not seem to allow them to be excluded. In such a case, a larger, more comprehensive study ought to be mounted to address, in part, any gender differences that may be present. Ultimately, the inclusion of women in medical research is a recognition that "[women's] liberation begins outside, with a succession of claims that this or that social good should be distributed for its own,

not for familial, reasons."[69]

The elderly. Until recently, the exclusion of older persons from clinical research was not seen as an ethical issue. (Indeed, it is rarely acknowledged that the exclusion of women from cardiovascular clinical trials may largely be due to the exclusion of older persons from those trials -- women develop heart disease later than men.[70]) Particularly within oncology, there is a growing recognition that older patients have too long been excluded from research participation.[71] Older persons are typically excluded from oncology research because they are thought to be more vulnerable to toxic effects from treatment.

Cancer clinical trials funded by the U.S. National Cancer Institute have excluded older patients for years.[72] For example, NSABP clinical trials examining chemotherapy in the treatment of stage II, node positive breast cancer excluded persons over the age of 70 from participation until 1981 (NSABP B-12 was the last protocol to have such an exclusion). In NSABP B-15 and subsequent protocols, the age exclusion was replaced with the criterion, "Patients must have a life expectancy of at least 10 years excluding their diagnosis of cancer." But there is no evidence that more older persons are being accrued to NSABP clinical trials. Indeed, since the change in the age criterion, the proportion of persons over age 60 years in the NSABP trials actually dropped.[73] Trimble and colleagues compared 1992 enrollment data for U.S. National Cancer Institute sponsored cooperative

group treatments trials with 1990 SEER incidence data by sex and cancer site.[74] For men and women over the age of sixty- five years, the differences between accrual and incidence -- direction in favor of underrepresentation in all groups -- were statistically significant for all cancer sites except prostate cancer. For persons over seventy-five years, accrual was significantly less than incidence in all groups.

The exclusion of older patients from medical research has led to a lack of information on proper treatment of a group of patients who carry the majority of the burden of the disease (age is the largest risk factor for the development of cancer). This lack of information, in turn, puts older patients at risk of undertreatment for their disease. Mor and colleagues reviewed the records of 1891 deceased cancer patients who had been associated with the National Hospice Study.[75] Controlling for stage and co-morbid disease, older patients were less likely to receive either chemotherapy or radiation therapy than younger cancer patients. Samet and colleagues reviewed data in the New Mexico Tumor registry on 22,899 cancer cases diagnosed between 1969 and 1982.[76] For most cancer types, the proportion of cases who received potentially curative therapy declined with patient age.

Studies that specifically looked at the treatment of older persons with breast cancer show similar findings. Allen and colleagues report on the treatment received by 1795 women with breast cancer referred to the Duke University

Medical Center between 1970 and 1984.[77] They found that older patients were more likely to receive surgery as the sole therapy for their disease. Furthermore, older women with nodal involvement were less likely to receive adjuvant chemotherapy than younger women (27% *versus* 60%). Chu and colleagues studied the care of 1680 women with breast cancer treated in 1982 at one of 17 hospitals associated with the Community Hospital Oncology Program.[78] Older women with breast cancer seemed to be broadly disadvantaged in the care they received. Older women received fewer diagnostic tests (biopsy, mammography), they were less likely to be seen by consultants, and they were (for all stages of disease) less likely to receive chemotherapy.

So then, older persons with cancer have been excluded from many cancer studies and treatment disparities exist between older and younger cancer patients. As persons get older they come to be more dependent on other family members, but this does not necessarily mean that they are less likely to benefit from cancer treatments; "dependency" and "vulnerability to toxic effects of chemotherapy" are not equivalent. Older patients have been excluded from trials for years on the presumption that they are more susceptible to the toxic effects of therapy, presumptions which may have been fueled by their often dependent role within the family, and yet "the assumptions upon which those policies were based have not been substantiated by empirical scrutiny."[79] Indeed, recent studies seem to indicate

that, at least with regard to some chemotherapy regimens used in breast cancer, older patients are not more vulnerable to toxic effects.[80] One might productively ask why such presumptions went unchallenged for so long.

New areas for ethical attention

The unjust inclusion of groups in research is driven by widely-held beliefs about the worth of members of such groups; the unjust exclusion of groups is driven by socially-constructed notions of deviation from the norm or vulnerability. Women are excluded from medical research because their reproductive physiology is a "complicating factor" and the potential for toxicity to fetuses. The elderly are blocked from participation because they are though to be more vulnerable to harm than younger people. In both of the cases we examined, the exclusion of these groups went unchallenged despite the lack of a convincing scientific justification. One worthy path of inquiry might examine other groups that are routinely excluded from studies and ask: Why is the group excluded? Is there any evidence to support such a claim?

For example, persons with a history of drug or alcohol abuse are often excluded from clinical trials. Presumably (specific reasons for individual exclusions are rarely given in clinical trial protocols), such persons are barred from participation because they are thought to be unreliable: they are unlikely to comply

with treatments and required follow-up visits. When Hughes reviewed the available literature to substantiate this claim, however, he found that there was no empirical basis for it whatsoever.[81] Indeed, compliance may be adequate even in persons with *active* drug abuse (let alone in those with a mere history of abuse). Harrison and colleagues report on comprehension and compliance in active intravenous drug users (IDUs; n=39) and non-IDUs (n=32) in a phase II trial of an HIV vaccine.[82] Both groups demonstrated satisfactory comprehension of consent information (out of a total possible score of 18 [mean ± SD]: IDUs scored 15.1 ± 2.0; non-IDUs scored 15.4 ± 1.8) and compliance in the two groups was similar (proportion receiving the fourth injection in series: IDUs, 67%; non-IDUs, 72%).

Other groups are (perhaps) also unjustly excluded from medical research. Otherwise healthy persons who are sero-positive for HIV are often excluded from studies. But are HIV-positive individuals really more likely to suffer adverse events or intercurrent illnesses during the course of a research study? Persons with a history of mental illness are sometimes excluded from research participation. Are they, as a group, any more likely to be incompetent to give informed consent? My guess is that further research will answer these questions in the negative.

Privately-funded research

One final question must be addressed, "Do these considerations only apply

to publicly-funded research and is privately-funded research exempt?" The question itself stems from the fact that the community neither has an obligation to provide for all needs nor is there a requirement to provide for each need to an unlimited extent. As we have indicated, needs must be prioritized by the community and only those which the community believes to be essential to its conception of the common life must be provided for.[83] Even the degree to which essential needs must be provided is subject to political limitation: labelling health care as an essential need does not imply that community members must be provided an unlimited amount of this good; the community has a right to set limits to its provision. The point is, says Walzer, that "[o]nce the community undertakes to provide some needed good, it must provide it to all the members who need it in proportion to their needs."[84] But if research is privately funded, i.e., not communally provided for, isn't it immune to such considerations?

Interestingly enough, recent changes to U.S. research guidelines might be construed as being in accord with an affirmative answer to this question. The recent *NIH Guidelines on the Inclusion of Women and Minorities as Subjects of Clinical Research* only apply to NIH-funded (i.e., publicly-funded) research. While 1993 changes to FDA regulations (largely governing privately-funded research) removed barriers to the enrollment of women of "reproductive potential" in clinical trials, the regulations stop short of *requiring* (as the *NIH Guidelines* do)

adequate representation of women in studies.

One line of argumentation against an exemption for privately-funded research might go as follows: Even if privately-funded research is exempt from the considerations of justice which we have outlined, little, if any, research today is wholly privately-funded. Pharmaceutical companies are given subsidies by the community to support medical research. For example, expenditures for research and development are not taxed; thereby, pharmaceutical companies are both encouraged to invest in the development of new medical treatments and pay substantially less tax on profits than they otherwise would. Indirectly, pharmaceutical companies depend on communal structures in order to do business. The drug licensure process is established by the community in order to ensure that new treatments are both safe and effective. While private companies bear the cost of research in support of new drug applications, the community bears the cost of the drug-approval bureaucracy itself. Thus, the line of argument goes, no (or little) current research in Canada or the U.S. is truly privately funded and, hence, no (or few) research studies are exempt from the considerations of justice as outlined.

While such an argument may appear to be functional, it is without elegance: it would still exempt *truly* private research. More importantly, though, I believe it misconstrues the problem. The argument proceeds as if research itself were the good to be distributed. It is not. Medical research, in our society at least, is

subservient to the social good of medical care. Research must, therefore, further the provision of medical care in accord with need and upholding the underlying equality of community membership. Once medical care is provided for communally, research in support of such care must ensure that results of clinical studies are applicable across the breadth of the affected community. The source of funding for such research is not relevant to considerations of justice.

But what of medical care for which the community does not provide? How shall we understand such care and the research that supports it? In some cases, treatments are not provided for because the community does not recognize the conditions which they ameliorate as diseases: for example, certain types of cosmetic surgery. In other cases, a community may sets limits on the amount of care to which an individual is entitled: for example, heart-lung transplantation may be deemed, given the scarcity of communal resources, an unacceptable expense. These two categories of non-communally provided treatments are best understood not as medical care, but rather as commodities.

If such treatments are commodities, it follows that community members do not have a right to them. Walzer says, "Beyond whatever is communally provided, no one is entitled to this or that useful or pleasing object."[85] People with enough money to pay for a facelift are entitled to it, those who can't pay for it, are not.

[T]here is no such thing as a maldistribution of consumer goods. It just

> doesn't matter, from the standpoint of complex equality, that you have a yacht and I don't, or that the sound system of her hi-fi set is greatly superior to his, or that we buy our rugs from Sears Roebuck and they get theirs from the Orient. People will focus on such matters, or not: that is a question of culture, not of distributive justice. So long as yachts and hi-fi sets and rugs have only use value and individualized symbolic value, their unequal distribution doesn't matter.[86]

Private companies doing research on such commodity-treatments would indeed have different rules to follow. Since such research operates within the sphere of money and commodities, and not in the sphere of security and welfare, economic justifications for the exclusion of subjects from research studies are allowed. A company may exclude righteously, for example, women from research studies examining a commodity-treatment if it can prove that it is cost-effective to do so. But issues of domination apply to the sphere of money and commodities, as to any sphere. Companies may not exclude persons from such research (or, for that matter, may not include them) by virtue of their standing in another sphere (and without appeal to cost-effectiveness).

Thus, when research supports medical care which is communally provided for, it must, irrespective of the source of funding, obey the justice considerations particular to the sphere of security and welfare. Treatments that the community has

decided not to provide for are not medical care but commodities. Research furthering such commodity-treatments must follow the rules particular to the sphere of money and commodities. Domination is, of course, not allowed in either sphere. Walzer says of medical care: "Needed goods are not commodities."[87] Our argument acknowledges this fact and adds: Neither are commodities needed goods.

Conclusion

Walzer's account of the moral world can be characterized by two premises. First, no meaningful or useful account of justice can be derived from abstract considerations. Ideas about justice flow from communally shared understandings, an understanding of the common life. Second, rules of justice are not universal in their application, they operate within spheres of justice, circumscribed domains of legitimate operation. The approach we take to justice issues must, therefore, be sensitive to history and context. Furthermore, and most important here, two types of injustice follow: violations of distributive rules within the relevant sphere of justice, and violations of the boundaries of the sphere (domination). In our account of justice in subject selection for medical research we have encountered both categories of injustice.

To date, accounts of justice and the selection of subjects for medical research have not provided us with an account of a just selection procedure. We

have argued that a just selection procedure is governed by the rule: *eligibility criteria must select subjects solely in accord with the exigencies of medical care and science*. Since the knowledge gained from research furthers the social good of medical care, such knowledge must be applicable to all members of the community. Thus, within the sphere of medical research the issue of generalizability is predominant. We have argued that if the community is to be served, eligibility criteria ought to be minimized and necessary exclusions ought to be explicitly justified.

Ultimately, current policy regarding the inclusion of women and members of minority groups in medical research are too narrowly construed. Once the unjust inclusion and exclusion of subjects from research is seen as boundary crossing, as domination, it is clear that many groups are potentially affected: not only women and minorities, but also the elderly, the impoverished, the undereducated and the politically disempowered. If justice is to be achieved, all of these groups must be our concern. Fundamentally, justice in medical research will not be achieved by "bean counting;" rather our goal must be to eliminate the selection of subjects for research on the basis of their standing in other spheres, be it education, political power, or kinship and love. In short, our goal must be to eliminate domination.

So too, the current NIH regulations are unduly limited in their focus on publicly-funded research. If the provision of medical care is recognized as a social

good, then all research in support of this good, including that which is privately funded, is subject to the justice considerations as outlined in this paper. To be sure, research which relates to commodity-treatments that are not communally provided for, e.g., certain types of cosmetic surgery, are subject to different distributive rules. But, even within the realm of commodity-treatment research, domination, i.e., selecting subjects solely because of their standing in other spheres, is prohibited.

A recognition of these moral facts will, I think, take us a little closer to the complex egalitarian society envisioned by Walzer.

> This is the lively hope named by the word *equality*: no more bowing and scraping, fawning and toadying; no more fearful trembling; no more high-and-mightiness; no more masters, no more slaves. It is not a hope for the elimination of differences; we don't all have to be the same or have the same amounts of the same things. Men and women are one another's equals (for all important moral and political purposes) when no one possesses or controls the means of domination.[88]

Acknowledgements

Earlier versions of this paper were presented at the Center for Applied and Professional Ethics, University of Tennessee at Knoxville in April 1996, and at the Canadian Bioethics Society Conference, Montreal, Canada in October 1996. The paper was originally prepared for a seminar on theories of justice given by Professor Karen Lebacqz at McGill University in 1996. The author is grateful to Professors Lebacqz and Freedman for their helpful comments and advice given in the preparation of this paper.

References

1.The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, "The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research," *OPRR Reports* April 18, 1979: 1-8.

2. Charles Weijer, "Evolving Ethical Issues in The Selection Of Subjects for Clinical Research," *Cambridge Quarterly of Healthcare Ethics* 5, 1996: 334-345.

3.Robert J. Levine, *Ethics and Regulation of Clinical Research*, 2nd ed., New Haven: Yale University Press, 1988: pp. 67-93.

4.P.V. Cardon, F.W. Dommel, R.R. Truble, "Injuries to Research Subjects: A Survey of Investigators", *New England Journal of Medicine* 295, 1976: 650-654; C.J.D. Zarafonetis, P.A. Riley, P.W. Willis, et al.. "Clinically Significant Adverse Effects in a Phase One Testing Program", *Clinical Pharmacology and Therapeutics* 24, 1978: 127-132; D.J. McCann, J.R. Pettit, "A Report on Adverse Effects Insurance for Human Subjects", in: President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioural Research, *Compensating for Research Injuries: the Ethical and Legal Implications of Programs to Redress Injuries Caused By Biomedical and Behavioural Research (Appendices)*, Washington, D.C.: Government Printing Office, 1982; J.D. Arnold, "Incidence of Injury During Clinical Pharmacology Research and Indemnification

of Injured Research Subjects at The Quincy Research Center", in: President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioural Research, *Compensating for Research Injuries: the Ethical and Legal Implications of Programs to Redress Injuries Caused By Biomedical and Behavioural Research (Appendices)*, Washington, D.C.: Government Printing Office, 1982.

5.Robert J. Levine, "The Impact of HIV Infection on Society's Perception of Clinical Trials", *Kennedy Institute of Ethics Journal* 4, 1994: 93-98.

6.Robert J. Levine, "Recruitment and Retention of Women in Clinical Studies: Ethical Considerations", in: A.C. Mastroianni, R. Faden, D. Federman (eds.), *Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies. Volume II: Workshop and Commissioned Papers*, Washington: National Academy Press, 1994: pp. 57-64.

7.*ibid.*

8.R. Macklin, G. Friedland, "AIDS Research: The Ethics of Clinical Trials", *Law, Medicine and Health Care* 14, 1986: 273-280.

9.Rebecca Dresser, "Wanted Single, White Male for Medical Research", *Hastings Center Report* January/ February, 1992: 24-29.

10.R.B. Merkatz, R. Temple, S. Sobel, et al., "Women in Clinical Trials of New Drugs: A Change in Food and Drug Administration Policy", *New England Journal of Medicine* 329, 1993: 292-296.

11.Department of Health and Human Services, National Institutes of Health, "NIH Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research", *Federal Register* 59, 1994: 14508-14513.

12.These responsibilities are laid out in a letter to Institutional Review Board Chairs by Gary Ellis, Director, Office for the Protection from Research Risks, dated April 25, 1994.

13.Susan Sherwin, "Women in Clinical Studies: A Feminist View", in: A.C. Mastroianni, R. Faden, D. Federman (eds.), *Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies. Volume II: Workshop and Commissioned Papers*, Washington: National Academy Press, 1994: pp. 11-17.

14.D.A. DeBruin, "Justice and the Inclusion of Women in Clinical Studies: A Conceptual Framework", in: A.C. Mastroianni, R. Faden, D. Federman (eds.), *Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies. Volume II: Workshop and Commissioned Papers*, Washington: National Academy Press, 1994: pp. 127-150.

15.Michael Walzer, *Spheres of Justice: A Defence of Pluralism and Equality*, New York: Basic Books, 1983.

16.Walzer, *op. cit.*, pp. xii-xii.

17.Walzer, *op. cit.*, p. 5.

18.Walzer, *op. cit.*, p. 7.

19.Walzer, *op. cit.*, pp. 6-10.

20.Walzer, *op. cit.*, p.10.

21.Walzer, *op. cit.*, p. 19.

22.Walzer, *op. cit.*, p. 20.

23.Walzer, *op. cit.*, pp. 10-11.

24.Walzer, *op. cit.*, p. 26.

25.Walzer, *op. cit.*, p. 17.

26.Walzer, *op. cit.*, p. 64.

27.Walzer, *op. cit.*, p. 87.

28.Walzer, *op. cit.*, p. 84.

29.Walzer, *op. cit.*, p. 88.

30.*ibid.*

31.James T. Patterson, *The Dread Disease: Cancer and Modern American Culture*, Cambridge, MA: Harvard University Press, 1987.

32.Patterson, *op. cit.*, p. 21.

33.Patterson, *op. cit.*, p. 47.

34.Patterson, *op. cit.*, p. 124.

35."National Cancer Institute Act of 1937", 78, 1987: 1017-1020.

36.Committee on Labor and Public Welfare, United States Senate, *National Program for the Conquest of Cancer: Report of the National Panel of Consultants on the Conquest of Cancer*, Washington: U.S. Government Printing Office, 1971: page 1; U.S. Department of Health, Education, and Welfare, Public Health

Service, National Institutes of Health, *National Cancer Program: The Strategic Plan*, Washington: DHEW Publication No. (NIH) 74-569, 1973: p. II-1, II-2.

37.American Medical Association Council on Scientific Affairs, "Viability of Cancer Clinical Research: Patient Accrual, Coverage, and Reimbursement", *Journal of the National Cancer Institute* 83, 1991: 254-259; R.E. Wittes, M.A. Friedman, "Editorial: Accrual to Clinical Trials", *Journal of the National Cancer Institute* 80, 1988: 884-885.

38.M.A. Friedman, D.F. Cain, "National Cancer Institute Sponsored Cooperative Clinical Trials", *Cancer* 65, 1990: 2376-2382.

39.Walzer, *op. cit.*, pp. 92-94.

40.G.B. Kolata, "The Death of a Research Subject", *Hastings Center Report* 10(4), 1980: 5-6; A. Darragh, M. Kenny, R. Lambe, I. Brick I, "Sudden Death of a Volunteer, *Lancet* 1, 1985: 93-94.

41.*Belmont Report*, *op. cit.*.

42.Walzer, *op. cit.*, p. 94.

43.*ibid.*

44.Walzer, *op. cit.*, p. 84.

45.Abraham Fuks, Charles Weijer, Benjamin Freedman, Stanley Shapiro, Myriam Skrutkowska, Amina Riaz, "A Study in Contrasts: Eligibility Criteria in a Twenty-year Sample of NSABP and POG Clinical Trials", under submission.

46.*ibid.*

47.Colin B. Begg, Paul F. Engstrom, "Eligibility and Extrapolation in Cancer Clinical Trials", *Journal of Clinical Oncology* 5, 1987: 962-968.

48.Benjamin Freedman, "Unethical research", in: W.T. Reich, ed., *Encyclopedia of Bioethics*, New York: Simon and Schuster MacMillan, 1995: pp. 2258-2261.

49.G.J. Annas, M.A., Grodin (eds.), *The Nazi Doctors and the Nuremberg Code: Human Rights in Human Experimentation*, New York: Oxford University Press, 1992.

50.R.L. Berger, "Nazi Science -- The Dachau Hypothermia Experiments, *New England Journal of Medicine* 322, 1990: 1462-1464.

51.J.H. Jones, *Bad Blood: The Tuskegee Syphilis Experiment*, New York: Free Press, 1993.

52.A.C. Mastroianni, R. Faden, D. Federman, *Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies. Volume I*, Washington: National Academy Press, 1994: pp. 75-83.

53.A.C. Mastroianni, R. Faden, D. Federman, *Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies. Volume I*, Washington: National Academy Press, 1994: pp. 108-127.

54.Food and Drug Administration, *General Consideration for the Clinical Evaluation of New Drugs*, Washington: Government Printing Office, 1977. (Publication no. HEW (FDA) 77-3040.)

55.General Accounting Office, *Women's Health: FDA Needs to Ensure More*

Study of Gender Differences in Prescription Drug Testing, Washington: Government Printing Office, 1992. (Publication no. GAO/HRD-93-17.)

56.C.E. Bird, "Women's Representation as Subjects in Clinical Studies: A Pilot Study of Research Published in JAMA in 1990 and 1992", in: A.C. Mastroianni, R. Faden, D. Federman (eds.), *Women and Health Research: Ethics and Legal Issues of Including Women in Clinical Studies. Volume 2: Workshop and Commissioned Papers*, Washington: National Academy Press, 1994: pp. 151-173.

57.R.S. Ungerleider, M.A. Friedman, "Sex, Trials, and Datatapes", *Journal of the National Cancer Institute* 83, 1991: 6-7.

58.A.C. Mastroianni, R. Faden, D. Federman, *Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies. Volume I*, Washington: National Academy Press, 1994: pp. 36-74.

59.J.Z. Ayanian, A.M. Epstein, "Differences in the Use of Procedures Between Women and Men Hospitalized for Coronary Artery Disease", *New England Journal of Medicine* 325, 1991: 221-225.

60.R.M. Steingart, M. Packer, P. Hamm, M.E. Coglianese, B. Gersh, E.M. Geltman, et al., "Sex Differences in the Management of Coronary Artery Disease", *New England Journal of Medicine* 325, 1991: 226-230.

61.Council on Ethical and Judicial Affairs, American Medical Association, "Gender Disparities in Clinical Decision Making", *JAMA* 266, 1991: 559-562.

62.Mastroianni, *op. cit.*, pp. 108-127, 175-202.

63.Walzer, *op. cit.*, p. 240.

64.Mastroianni, *op. cit.*, p. 113.

65.W.B. Patterson, E.J. Emanuel, "The Eligibility of Women for Clinical Research Trials", *Journal of Clinical Oncology* 13, 1995: 293-299.

66.*ibid.*

67.B. Healy, "The Yentl Syndrome", *New England Journal of Medicine* 325, 1991: 274-275.

68.Walzer, *op. cit.*, p. 84.

69.Walzer, *op cit.*, p. 240.

70.J.H. Gurwitz, N.F. Col, J. Avorn , "The Exclusion of the Elderly and Women from Clinical Trials in Acute Myocardial Infarction", *JAMA* 268, 1992: 1417-1422.

71.B.J. Kennedy, "Needed: Clinical Trials for Older Patients", *Journal of Clinical Oncology* 9, 1991: 718-720.

72.C. Begg, P. Carbone, "Clinical Trials and Drug Toxicity in the Elderly: The Experience of the Eastern Cooperative Oncology Group", *Cancer* 52, 1983: 1986.

73.Fuks, *op. cit.*.

74.E.L. Trimble, C.C. Carter, D. Cain, B. Freidlin, R.S. Ungerleider, M.A. Friedman, Representation of Older Patients in Cancer Treatments Trials", *Cancer* 74 supp., 1994: 2208-2214.

75.V. Mor, S. Masterson-Allen, R.J. Goldberg, F.J. Cummings, A.S. Glicksman,

M.D. Fretwell, "Relationship Between Age at Diagnosis and Treatments Received by Cancer Patients", *Journal of the American Geriatric Society* 33, 1985: 585-589.

76.J. Samet, W.C. Hunt, C. Key, C.G. Humble, J.S. Goodwin, "Choice of Cancer Therapy Varies With Age of Patient", *JAMA* 255, 1986: 3385-3390.

77.C. Allen, E.B. Cox, K.G. Manton, H.J. Cohen, "Breast Cancer in the Elderly: Current Patterns of Care", *Journal of the American Geriatric Society* 34, 1986: 637-642.

78.J. Chu, P. Diehr, P. Feigl, G. Glaefke, C. Begg, A. Glicksman, L. Ford, "The Effect of Age on the Care of Women with Breast Cancer in Community Hospitals", *Journal of Gerontology* 42, 1987: 185-190.

79.Mor, *op. cit.*

80.S. Giovanazzi-Bannon, A. Rademaker, G. Lai, A.B. Benson, "Treatment Tolerance of Elderly Cancer Patients Entered onto Phase II Clinical Trials: An Illinois Cancer Center Study, *Journal of Clinical Oncology* 12, 1994: 2447-2452; K. Christman, H.B. Muss, L.D. Case, V. Stanley, "Chemotherapy of Metastatic Breast Cancer in the Elderly: The Piedmont Oncology Association Experience, *JAMA* 268, 1992: 57-62.

81.J.R. Hughes, "Exclusion of "Noncompliant" Individuals from Clinical Trials", *Controlled Clinical Trials* 14, 1993: 176-177.

82.K. Harrison, D. Vlahov, K. Jones, K. Charron, M.L. Clements, "Medical Eligibility, Comprehension of the Consent Process, and Retention of Injection

Drug Users Recruited for an HIV Vaccine Trial", *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 10, 1995: 386-390.

83.Walzer, *op. cit.*, p. 66.

84.Walzer, *op. cit.*, p. 75.

85.Walzer, *op. cit.*, p. 104.

86.Walzer, *op. cit.*, p. 108.

87.Walzer, *op. cit.*, p. 90.

88.Walzer, *op. cit.*, p. xiii.

Chapter 6:

Discussion:

Explanatory *versus* pragmatic approaches to selecting subjects for research participation.

Introduction

As stated in the introductory chapter, the purpose of this thesis is to examine aspects of one area of overlap between science and ethics, namely, selection procedures for participation in clinical research. Such procedures, formalized in eligibility criteria, were examined at a number of points in the genesis and dissemination of medical knowledge: clinical trial protocol, interpretation by researchers, and communication of study results. This final chapter has several purposes. First, to examine critically the overlapping scientific and ethical concerns that arise in the selection of subjects for clinical research participation. Second, to explain the methodology employed in the thesis and suggest further avenues for research. Third, and finally, to point out which aspects of the thesis represent original contributions to the literature.

Explanatory and pragmatic clinical trials

Our examination of NSABP and POG clinical trials in chapter 2 revealed two substantially different approaches to the selection of subjects for research participation. Following Schwarz and Lellouch's classic paper, we referred to the strategy illustrated by the NSABP trials as 'explanatory' and that followed by the POG trials as 'pragmatic.'[1] These two philosophies of clinical trials carry broad implications for the design, conduct and interpretation of clinical studies.

Explanatory clinical trials attempt to "discover whether a difference exists between two treatments which are specified by strict...definitions."[2] The aim of such trials is to deepen our *understanding* of a medical intervention under tightly controlled circumstances akin to those found in the laboratory. Pragmatic trials, on the other hand, "compare two treatments under the conditions in which they would be applied in practice...[they seek] to answer the question — which of the two treatments should we prefer?"[3] Pragmatic trials are oriented towards making a *decision* as to which treatment is to be preferred under clinical circumstances.

These two approaches to trial design imply different research questions and different methods to answer these questions. Schwarz and Lellouch give the example of a new drug which may sensitize tumor to the effects of radiotherapy. Normally the drug would be administered for thirty days prior to the first course of radiotherapy. An explanatory approach asks, Does the drug have the biological effect claimed? In other words, Is treatment with the drug followed by radiation therapy superior to no treatment for thirty days followed by radiation therapy (i.e., delayed therapy). The design for this explanatory study is illustrated in figure 1. A pragmatic approach is less concerned with biological effect than which treatment is to be preferred in the clinic, drug followed by radiotherapy *or* immediate radiotherapy. The design for this pragmatic study is also seen in figure 1. Explanatory and pragmatic approaches have implications for aspects of trials other

than design, including dosage of study drugs (equimolecular doses for comparison [explanatory] *versus* optimal doses for comparison [pragmatic]), choice of study endpoint (tumor regression *versus* survival), dealing with patients who withdraw from the study (excluding them from the analysis *versus* including them, i.e., an intention-to-treat analysis), and, the selection of subjects (carefully restricted set of participants *versus* an all-comers approach).

Selection procedures for study participation define who is suitable to participate in the study (and who is needed to answer the question at hand). Since an explanatory trial seeks to establish whether a biological effect exists, it will "be done on a relatively arbitrary population which is well adapted to the problem at hand, homogeneous and with low withdrawal rate."[4] Such studies will often *include* only patients who are deemed most likely to respond to the treatment and *exclude* patients who are thought likely to experience adverse effects. The proportion of the patient population eligible for such a study may be very low indeed. A pragmatic trial, aiming to decide which treatment ought to be used in practice, takes an all-comers stance with regard to subject selection.

> To enable the results to be extrapolated to a defined population of patients, the trial should be carried out on a properly representative sample of this population. This counsel of perfection is rarely followed, but the patients chosen for the trial must represent as far as

> possible the population to which the results are to be extrapolated. Extrapolation will be the more justifiable if the trial can rest on a broad range of sampling — this is one reason for undertaking collaborative trials at several centers.[5]

What are the consequences of explanatory and fastidious approaches for the interpretation of trial results? Explanatory trials only have immediate pragmatic implications when the new treatment is proved no better than the control treatment (this assumes, of course, that the study had adequate power "to detect" a clinically important effect). For example, consider the study conducted on a highly select group of patients: If the new treatment is proven effective, one still has no information on the therapeutic index of the treatment in the broader patient population. If the treatment is proven ineffective, it has failed under the most favorable conditions, and *a fortiori* it will be ineffective if tested in the patient population as a whole (i.e., under less favorable conditions). In short, only "negative" explanatory trials have immediate pragmatic implications.

Pragmatic trials have explanatory value only when the new treatment is proven superior to the control treatment. Consider the study performed on a heterogeneous population of patients: If the treatment is not found to be effective under such circumstances, it may yet be found to be effective for a more favorably defined subgroup of patients. Only if the treatment is found to be superior to the

comparison treatment, has the biological question been answered. Thus, only "positive" pragmatic trials have explanatory value.

In some circumstances, there is agreement as to whether an explanatory or pragmatic approach to trial design is appropriate. For example, phase II cancer trials attempt to establish whether a new anti-cancer treatment has an effect in a particular disease setting, and therefore, such trials utilize explanatory designs. Disagreement exists as to the preferred approach for phase III clinical trials. The disagreement occurs on a fundamental, even philosophical level. Feinstein characterizes the two camps of trial designers as follows:

> The proponents of one viewpoint [pragmatic] usually want the trials to answer pragmatic questions in clinical management. For this purpose, the plans would incorporate the heterogeneity, occasional or frequent ambiguity, and other 'messy' aspects of ordinary clinical practice. The advocates of the opposing viewpoint [explanatory] fear that this strategy will yield a 'messy' answer. They prefer a 'clean' arrangement, using homogeneous groups, reducing or eliminating ambiguity, and avoiding the spectre of biased results.[6]

As suggested above, this controversy results in two distinct approaches to defining patient populations for randomized clinical trials.

> A fastidious [explanatory] designer will want to test a relatively

> homogeneous group of patients and may therefore 'purify' the eligible candidates by including only people of one gender and race, within a limited age span, with no coexisting other diseases or treatments with medications, who have been checked for their willingness to cooperate well with the requirements of the experimental protocol...The pragmatic designer , however, will complain that the 'pure' results are often useless for practical application to the heterogeneous spectrum of cases and the many 'impurities' that are encountered in clinical reality.[7]

How can we account for such deep divisions in the approach to the design of phase III clinical trials? Only an incomplete answer can be given to this intriguing question. It seems that explanatory trials may be more common in North America and pragmatic trials more common on the other side of the Atlantic Ocean. In the comprehensive review of empirical literature of trial enrollment presented in my Master's thesis, a substantial discrepancy was observed between North-American trials and European studies.[8] North American studies (n=9) excluded 55% of potential subjects (patients with the appropriate type and stage of disease) whereas European studies (n=9) excluded only 36% of subjects. Some of the European clinical trials used very few eligibility criteria indeed. Anderson reports on three trials of the Danish Breast Cancer Cooperative Group that used

only three selection criteria: "operable breast cancer with no metastises," "no medical contraindication to the study treatments," and "age less than 70."[9] (Recall that the most recent NSABP study we reported on contained 44 eligibility criteria).

One can only speculate as to how the trans-Atlantic difference in trialing came about. Influential authors (and teachers) like England's Richard Peto[10] and America's Richard Simon[11] likely had some influence on the development of trial philosophies in the two continents. Payer has described substantial differences in the practice of medicine in the United States and a number of European countries which seem to be culturally driven.[12] It is conceivable that the acceptance of one trial philosophy versus another may similarly have roots in cultural factors. One's affinity for explanatory or pragmatic trials may even be viewed as containing an aesthetic component: What counts as an elegant experiment?

Against this backdrop, let us consider the scientific arguments made in favor of explanatory or pragmatic approaches to subject selection for clinical trials. What motivates the "explanatory trialist?" Can the "pragmatic trialist" respond adequately to the explanatory position?

Arguments for an explanatory approach to subject selection

> ...for it was realized that no two patients have an identical form of the disease and it was desired to eliminate as many of the obvious

> variations as possible. This planning...is a fundamental feature of the successful trial. To start out upon a trial with all and sundry included, and with the hope that the results can be sorted out statistically in the end is to court disaster.[13]

Including patients of differing prognosis (i.e., a heterogeneous study population) will increase variability in a study and, hence, decrease its power. (The arguments in favor of explanatory criteria for trial eligibility are summarized in table 1.) The randomized clinical trial originated relatively recently, within the last half century, and represents a major scientific advance in medicine: the relative safety and efficacy of novel medical treatments can be reliably assessed. The methodology for clinical trials was developed in other areas of science, primarily agriculture, and many of the features of "wet-bench experiments" were preserved with its introduction into medicine. A fundamental principle of the "wet-bench experiment" is to control every variable except the variable of interest, i.e., treatment. With regard to subject selection, then, a homogeneous study population, that is one in which prognosis is uniform, is regarded as ideal.

It is only when extraneous variables are tightly controlled, and thus, in part, when the patient population in a study is highly select, that we give a new treatment the best chance of being proven effective. If patients with differing prognoses are included into a clinical trial, the variability — statistical "noise" —

in the study is increased. The true effect of the study intervention — the "signal" — will tend to be lost as "noise" in a study increases. Gail argues that these facts demand that eligibility criteria select as homogeneous a group as possible:

> These restrictions tend to produce a fairly homogeneous study population so that the effects of treatment are more easily discernable against the background variability in response. If no such restrictions were imposed, the effect of treatment could be lost in the tremendous variability of response arising from mixed stages of disease and cell types, which are often more powerful determinants of outcome that is the treatment.[14]

In statistical parlance, if the variability in a study is increased, given the intervention has some actual effect, the probability of the study concluding that the treatment is effective (power) diminishes.[15] In short, without a carefully selected patient population, the chance of a study coming to a falsely negative conclusion is increased.

Studies involving homogeneous groups of patients are more efficient (in terms of "residual sample size requirements"). The second argument is derivable from the first. Two patient-related factors which contribute to the power of a study are (1) risk of the event of interest (death, recurrence, etc.) and (2) response to treatment. If the number of events is high and the treatment effect is large, then the power of the study will be high. It follows, then, that, for a fixed sample size, high-

risk, high-responder patients will contribute relatively more information to a trial than low-risk, low-responder patients. Thus, (again, for a fixed sample size) it is more efficient to add high-risk, high-responder patients — a highly selective population — to a clinical trial than members of other patient subgroups.

Sackett makes this point, and his comments are worth quoting *in extenso*:

> There is a definite advantage in admitting study patients who are both at high risk of an event (to pull event rates away from zero, where it requires very large numbers of study patients to show risk reductions even down to zero) and are highly responsive to the test therapy (thereby producing the largest differences in event rates between experimental and control patients). Similarly, we would like to exclude patients from our trial who are both at low risk of an event and unlikely to respond to the test therapy; indeed, the addition of such patients increases rather than decreases the residual sample size requirement for demonstrating a statistically significant between-group difference...The foregoing considerations suggest that investigators should, in the design stages of their trials, estimate the likely risk and responsiveness of various subgroups of potential study patients and establish inclusion/exclusion criteria that will limit entry to the high-risk, high-response subset...In summary, several

> limitations to the generalizability of the results of a trial are inevitable and, indeed, essential to the efficient and unambiguous demonstration of efficacy.[16]

Clearly, Sackett's argument begins with the premise that sample size is fixed, or at least trial resources are limited. (An assumption that many will find reasonable in this day and age). Given this, How to maximize the chance that an effective treatment will be found to be so by a study? Sackett's answer: study a select group of patients. Given the origins of the randomized clinical trial, an agricultural example may be appropriate. If one has a batch of seed that may or may not be too old to germinate, an efficient approach to answer the question is to plant it in the most fertile soil under optimal growing conditions. Why? If the seed will grow anywhere, it will grow there.

Within a heterogeneous study population, qualitative differences among subgroups may cancel one another out, thus obscuring the treatment's "true" effect. Both of the arguments cited above are linked by the concern that when patients with varying prognoses are included in a study, variance increases. In a broad patient population, not all of the patients will be high-responders; there is diversity with regard to treatment response. When differing groups of patients experience varying degrees of response to a treatment, but the treatment effect is all in the same direction, this is referred to as "quantitative interaction."

Quantitative interaction is a source of error within a study, but one that can be compensated for by increasing the study sample size.

More problematic is "qualitative interaction": subgroups of patients in whom the *direction* of treatment effect is different. Consider a study testing a new anti-cancer treatment in which half of the patients, on average, do better with the treatment, but the other half — perhaps due to age or some other factor — do substantially worse. Despite the fact that the treatment is effective in one large subgroup, in the overall analysis, the two groups will tend to cancel one another out leading to an overall conclusion of "no difference." Increasing the sample size will not necessarily compensate for this source of error. Simon makes this point repeatedly:[17] "In a study with broad eligibility requirements, a conclusion of no difference between the treatments may result from a positive effect in one subset being canceled by a negative effect in another..."[18]

Of course, if it was known *ab initio* that a subgroup of patients was likely to be harmed by a treatment (i.e., experience a negative outcome), these patients would be excluded by any responsible trialist — whatever the trial philosophy he or she embraces — from study participation. The above argument refers to *unidentified* subgroups embedded within the patient population. The assertion is that qualitative interactions will be less common when the patient population for a study is homogeneous rather than heterogeneous.

In a heterogeneous study population, there are insufficient numbers in individual subgroups to do a meaningful statistical analysis. But if subgroups of patients may vary in their response to treatment, couldn't we figure that out with subgroup analyses? In theory at least, the statistical analysis could examine treatment effect in the study according to various prognostic factors (age, comorbidity, etc.), thereby arriving at more precise estimates of treatment effect. The problem with this approach is that individual subgroups within the study may contain too few subjects to allow for a meaningful (read: adequate power) subgroup analysis. Sylvester articulates the problem as follows:

> [T]he patient population should be reasonably homogeneous so that all patients have a similar type of disease and prognosis. One should avoid including in the trial small subgroups of patients who have a potentially different prognosis from the others. There will not be enough of these patients to analyze them separately and including them in the analysis may weaken the overall treatment comparison.[19]

Once again, assuming a fixed sample size, only a homogeneous study sample seems to get around this problem.

Studies with broad inclusion criteria (and, hence, numerous subgroups of potential interest) may lead to misleading multiple subgroup analyses. This point is related to the previous concern. If a study includes a heterogeneous group of

patients, multiple subgroup analyses may be required. If the sample size in each group tested is inadequate, or if multiple tests are done but multiple testing is not corrected for, misleading results may result. Simon:

> Some statisticians advise that the eligibility criteria can be very broad because subset analyses can always be performed later. This approach has certain risks, however: misleading conclusions may result from multiple subset analyses, and one must be careful to plan the study so that adequate numbers of patients within each major subset are available for separate analysis.[20]

As with some of the above arguments, this argument suggests a preference for strict eligibility criteria if the sample size of the trial is fixed.

In summary, a variety of concerns have been put forward in the literature regarding the hazards and impracticality associated with heterogeneous groups of patients in randomized clinical trials. Including patients of differing prognosis may increase variability and decrease power; including only very select patients may, therefore, be more efficient; qualitative differences between patient subgroups may obscure beneficial effects of a treatment; subgroup analyses in many cases cannot be done without many more patients in the study; and, if they are done, they may be misleading. These arguments have convinced many that despite whatever drawbacks an explanatory approach to trial design may have, it is to be preferred

to a pragmatic approach.

Arguments for a pragmatic approach to subject selection

> The criteria for a good trial are fairly straightforward: ask an important question and answer it reliably. The importance of the question depends to a large extent on its clinical relevance. It is obvious that the more widely applicable are the results of a clinical trial, the more relevant and valuable are those results.[21]

A truly homogeneous patient population cannot be defined; patient-to-patient variability is the largest source of variation. (A summary of arguments for pragmatic approaches to patient selection is found in table 1.) Trial designers who advocate an explanatory approach to patient selection argue that restrictive eligibility criteria are necessary to define a homogeneous patient population. Trial pragmatists doubt whether added criteria actually define a population that is substantially more homogeneous. Once patients have been selected for participation in a clinical trial according to the most minimal criteria — type and stage of disease, no absolute contraindication to study treatment — few important prognostic factors remain. Whether or not one tries to further restrict the study population, patients in a clinical trial are a heterogeneous group — they vary in their response to medical treatments. Two patients — matched for age, cancer

type, stage of disease, other comorbid conditions, and treatment — may have substantially different outcomes: one may be cured while the other dies of rapidly progressive disease. Without good additional predictive factors, identified and implemented *a priori*, added criteria will not make a study population substantially more homogeneous. According to Begg and Engstrom

> [H]omogeneity is an ideal that is not even closely approximated in the clinical setting where patient heterogeneity is substantial even in narrowly restricted studies, especially with regard to prognosis, so that the between-patient variation is always large relative to the anticipated treatment effect.[22]

If we accept that patient-to-patient variation is large in studies (i.e., there are no good additional predictive factors), what follows? Yusuf concludes that patients need neither be screened intensively with restrictive eligibility criteria, nor, for that matter, precisely characterized at the beginning of a trial.[23] The variation in response to treatment can, according to the trial pragmatists, only be compensated for by enrolling large numbers of patients into trials, so that studies have adequate precision. Yusuf, Collins and Peto explain the necessity of this as follows:

> Clinicians are used to dealing with individual patients, and may feel that the results of large trials somehow deny the individuality of each

> patient. This is almost the opposite of the truth, for one of the main reasons why trials have to be large is just because patients are so different from one another. Two apparently similar patients may run entirely different clinical courses, one remaining stable and the other progressing rapidly to severe disability or early death. Consequently, it is only when really large groups of patients are compared that the proportions of truly good and bad prognosis patients in each can be relied on to be reasonably similar.[24]

A study's power is maximized by removing eligibility criteria and making the sample size larger, not further restricting the study population. Proponents of explanatory-trial philosophy argue that restricting the criteria for trial eligibility will increase the study's power. This would only be true if one could truly define a homogeneous study population, a notion cast into doubt by the above argument. Peto argues that *even if* a more homogeneous study population could be defined, one would be better off with a large, simple trial.[25] Sample size, according to Peto, is a more important determinant of a study's power than the efficiency contributions of individual patients. "The larger the size of the trial, the smaller the random error. Consequently, reliable overall results are more likely to emerge."[26]

It seems that very large trials may be necessary in order "to detect" reliably the effects of new treatments in cardiology and oncology (and, possibly, other

areas of medicine). Yusuf observes that medical advances most commonly come in small steps: the magnitude of treatment effects are, at best moderate — a 15% to 25% risk reduction in death or serious outcome.[27] Peto points out that the detection of moderate differences requires the stringent control of random and systematic error that only a large, randomized trial (or systematic overviews of a number of trials) can provide:

> If moderate differences in outcome are to be detected or refuted reliably, then the errors in comparative assessments of the effects of treatment must obviously be much less than the difference between a moderate but worthwhile effect, and an effect that is too small to bother with. This in turn implies that moderate biases cannot be tolerated, and moderate random errors cannot be tolerated: in practice, this implies the need for methods that involve negligible biases and very small random errors. The only way to guarantee very small random errors is to study really large numbers, and these can be achieved in two main ways: make individual studies large, and combine information from as many studies as possible. But it is not much use having very small random errors if there may well be moderate biases, so even the very large sizes of some non-randomized analyses of medical records cannot guarantee

> statistically reliable results.[28]

In order for a study to have sufficient power, 10 000 to 20 000 average-risk patients or 3 000 to 5 000 high-risk patients are often required.

Trials with highly restrictive eligibility criteria may have difficulty accruing adequate numbers of patients. If trials need to be *very large* in order to detect reliably moderate treatment effects, then they also need to be *simple*. Unnecessary eligibility criteria unduly restrict the pool of patients available for study and, thereby, diminish enrollment rates. A great deal more might be gained by eliminating such criteria than by retaining them. Peto:

> Any obstacle to simplicity is an obstacle to large size, so it is worth making efforts to simplify the process of entering, treating and assessing patients. It is particularly necessary to simplify the entry of patients, for if this is made complicated then recruitment may be very seriously damaged.[29]

His concluding statement is, if somewhat an overstatement, though-provoking: "Most trials would be of much greater scientific value if they collected ten times less data, both at entry and during follow-up, and were therefore much larger."[30]

Of course, eligibility criteria are not merely a barrier to accrual in very large trials. We pointed out earlier that of eligibility criteria, physician factors and patient factors, eligibility criteria were the most important barrier to trial accrual.

Since many trials in oncology take longer to enroll their target population than planned, George urges that selection criteria be minimized in all randomized clinical trials:

> An obvious implication of restrictive eligibility criteria is that each criterion will reduce accrual. Since a major impediment to completing many trials successfully is a low accrual rate, impediments to accrual should be removed if at all possible.[31]

Added eligibility criteria increase trial complexity and costs; minimizing criteria reduce complexity and cost. As suggested in Peto's comments above, the costs of unduly restrictive eligibility criteria are not limited to trial enrollment. Unnecessary criteria burden trial personnel and resources.[32] Clinical investigators must spend more time with prospective subjects, additional tests must be ordered and interpreted, and added time must be spent filling out complex enrollment forms. Data managers must collect and store greater amounts of information on each trial participant. Trial resources are taxed by the fact that the cost per patient is increased by added tests and physician time.

We concluded above that minimizing selection procedures would enhance trial enrollment. We see here that this benefit stems from more than a larger pool of eligible patients, simplifying criteria makes it *easier for investigators* (less time, less tests, shorter forms) to enroll patients as well. Additionally, the cost per

patient is reduced and, therefore, even with fixed resources, trials can afford to be larger. Yusuf: "in general, wide eligibility criteria substantially simplifies screening and increases recruitment. This in turn reduces the effort and the cost per patient enrolled, thereby making large trials more affordable and feasible."[33]

Studies involving heterogeneous groups of patients are more efficient (in terms of accrual rate and time to completion). Explanatory trialists make the claim that narrowly-focused trials are more efficient: a homogeneous patient population minimizes the number of patients needed to answer a question reliably. But "residual sample size requirements" are not the only way of understanding efficiency. It asks: What is gained by adding to the trial this particular subject compared with adding another from the range of potential subjects? Rather than approaching efficiency on a per subject basis, we might ask instead, How long will it take to complete a trial with one policy of patient selection *versus* another?

Buyse asks precisely this question. He considers a hypothetical — but plausible — example in which a policy of including poor prognosis patients is compared with a policy of restricting trial entry to good prognosis patients. A broader patient selection policy is preferred (i.e., the trial is completed faster) as the proportion of poor prognosis patients in the population increases and as their rate of treatment response approaches that of the good prognosis group. Buyse concludes that over a broad range of assumptions, a more pragmatic trial will

answer the study question more quickly than an explanatory trial:

> [T]he model shows that broad trials are preferable to restricted trials in many situations ('preferable' being taken in the limited sense of 'shorter in duration'). One such situation is when there is no *a priori* reason to believe that the magnitude of the treatment effect is different in different subgroups of patients. In such a case, the best strategy (from a statistical standpoint) is to include all available patients in the clinical trial, whatever the prognosis. This situation occurs quite commonly...[34]

George also considers the problem from a theoretical perspective.[35] He points out that even if a larger sample size is required by a trial with a heterogeneous patient population, that requirement is likely to be more than offset by the increase in the rate of enrollment to the study. A broadly-inclusive trial will come to completion more quickly than a narrowly-focused study, particularly when added patients have a similar response to treatment and result in a substantial increase in study accrual.

Qualitative differences among subgroups are, generally speaking, uncommon. As pointed out above, unexpected qualitative differences can have serious ramifications for a clinical trial: differences in the direction of treatment effect between subgroups will tend to cancel one another out. But are such

interactions likely? Are they common? Not all patients are eligible for even the most pragmatic trial. Even selection procedures advocated by trial pragmatists require that subjects have the type and stage of disease of interest and no major contraindications to the study treatments. These criteria mirror factors used in clinical practice and, thus, do not limit the generalizability of trial results to clinical practice. The real question here is once patients with known contraindications have been excluded, are qualitative interactions likely? It seems the answer is, "No."

Yusuf believes that "the probability is low of reliably finding an *unanticipated* qualitative interaction (that is, differences in the direction of effect) in a trial that has already excluded those in whom treatment is clearly indicated or contraindicated."[36] In his paper, Yusuf reviews a studies of treatments for vascular disease which involves a wide variety of study populations (refer to table 2). Despite the heterogeneity of patients across studies, no evidence of qualitative interaction is found. (Indeed, for the most part, even the magnitude of treatment effect is similar across groups.) Surprisingly, even when large quantitative interactions or qualitative interactions were predicted on the basis of prior work — e.g., thrombolysis more than six hours post-myocardial infarction, beta-blockers in patients with heart failure — these interactions did not materialize in well done clinical trials. Peto reviews the experience with a number of very large trials and

meta-analyses with similar findings. Peto concludes that "for specific outcomes, the directions of the effect of treatment may be similar in many different categories of patient."[37]

Statistical techniques exist to adjust for important known covariates. What if important prognostic factors *are* known to exist? Must patients be excluded in these situations? Not necessarily. If the prognostic factors identify patients who will be treated with the same or similar treatments in clinical practice, then they probably ought to be included in a clinical study. The presence of important predictive factors can be accommodated in the study's design (e.g., stratified randomization) or analysis. Begg and Engstrom point out that:

> [W]e have available a wide range of statistical techniques that deal effectively with patient heterogeneity, both in the design of the study and at the analysis stage, although the selection of an adjustment technique, i.e., fully or partially stratified analysis *v.* covariate adjustment by regression-type model (e.g., Cox proportional hazards model), is a matter for informed judgement.[38]

The results of narrowly-focused studies may not be applicable to the patient population at large; trials with heterogeneous patient populations are more likely to be widely applicable in clinical practice. Perhaps the most important critique of explanatory trials is the fact that the results of such trials may not be widely

applicable in clinical practice. Leventhal observes that

> The results of clinical trials are most useful when their results can be generalized with reasonable confidence to the disease population at large. To know how generalizable results are likely to be, one must know how representative the patients studied are of all patients with the disease. One important indication of this is a numerical measure of how highly selected they are, i.e., how many of the patients seen with that diagnosis during the study period were entered in the trial. In addition it is important to know how representative the study group is in terms of patient characteristics...If the patients in a clinical trial are not representative of the entire patient population... the generalizability of the results to the entire patient population may be compromised.[39]

Indeed, the results of clinical trials are often criticized on the grounds that the study was too restrictive to be widely applicable in practice.[40]

We saw in chapter 2 that cancer clinical trials routinely exclude persons on the basis of age (or expected survival aside from the cancer diagnosis) or laboratory values. Such exclusions lead to uncertainties regarding the proper treatment of excluded groups in clinical practice. Ultimately, George believes, the presence of numerous criteria may diminish the *overall impact* of the study:

> An important goal of a clinical trial is to change clinical practice. However, if the eligibility criteria do not reasonably reflect the type of patient likely to be treated in general clinical practice, the result will be unpredictable. Clinicians are unlikely to go through a long eligibility checklist of dubious relevance and will have difficulty assessing whether the result obtained in relevant to their patients.[41]

The fact is that many persons currently excluded from cancer clinical trials *require and often will receive treatment with similar agents in practice*. The failure to include such groups in trial may lead to one of two equally undesirable outcomes. First, patients may not receive possibly effective treatments. In chapter 5 we argued that the exclusion of the elderly from cancer clinical trials predisposes older persons with cancer to under-treatment. Stenning agrees that the inclusion of older persons in trials is important:

> [M]any trials of therapy for colorectal cancer will include upper age limits. In trying to conduct a 'clean' trial, and to this end setting the limit such that it will preclude entry to patients with too great a chance of dying from an intercurrent condition, it would be easy to eliminate a large proportion of patients with colorectal cancer. An upper age limit of 70 years would exclude over a third of patients with the disease, and would give no information on treatment

tolerance in the older patient...[42]

Second, patients may receive treatment outside of trials which is of unproven value. Stenning discusses a clinical trial of chemotherapy for locally advanced cancer of the bladder in which subjects, in order to be eligible for participation, had to have a glomerular filtration rate (GFR; a measure of renal function) of at least 60 milliliters per minute.

> Some time after the launch of the trial, it was found that many patients were being excluded because of renal function requirements not being fulfilled. However, many of the patients were actually receiving chemotherapy outside of the trial, albeit at lower doses than those specified in the protocol, for example cisplatin doses of 70 mg/m^2 [compared with 100 mg/m^2 in the trial]. An alternative trial design might include the option of allowing patients with reduced renal function to be randomized, but receiving lower drug doses, rather than excluding them altogether.[43]

Including these patients in the trial would allow the risks and benefits of chemotherapy in this group to be evaluated.

Again, an agricultural example may be appropriate here. Recall our batch of seed which may or may not be too old to germinate. Let us imagine that the above suggested experiment has been carried out and the seed had an acceptable

germination rate under optimal conditions. A farmer considering a large purchase of such seed (perhaps it comes at a discounted price) is likely to refuse the purchase without further evidence. Before an informed and responsible decision can be made, she must know, "Will the seed germinate under the varied conditions that are found in *my* several fields?" The answer to this question requires that germination be assessed under heterogeneous soil conditions: in fields that have grown crops for years, in fields that remain fallow; in fields that receive ample rainfall, in those that are dry; in soil that is alkaline, in soil that is not. Only when assessed under this multiplicity of conditions (and only if it performs under these conditions) can the seed be said to be a good buy.

In summary, a variety of powerful arguments have contradicted claims made in favor of explanatory trials and extolled the virtues of pragmatic trials, studies with heterogeneous patient populations. These arguments included the following claims: beyond that achieved by minimal criteria, a homogeneous population is an unattainable ideal; very large trials are necessary "to detect" moderate but important treatment effects; in order for trials to be very large, they must be simple (i.e., minimize eligibility criteria); eliminating eligibility criteria not only improves enrollment rates, it reduces complexity and cost; simple trials (with heterogeneous populations) accrue patients faster thereby answering the study question more efficiently (faster); qualitative interactions are uncommon;

even if important covariates are known to exist, they can be adjusted for by design or analysis; and, most important, only the results of broad-based, inclusive studies are likely to be widely applicable in clinical practice. The weight of these arguments has convinced many that pragmatic approaches to the design of phase III clinical trials are to be preferred.

Where science and ethics meet

The choice of an explanatory or pragmatic approach to the design of a clinical trial has implications not only for science, but also for ethics. What is the basis for ethical concerns regarding the selection of subjects for research participation? What are the implications for the preference of an explanatory *versus* pragmatic approach to trial design?

In the first chapter, we discussed the ethical framework presented in the *Belmont Report.* In this document, the members of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research articulate three principles to guide the conduct of research: respect for persons, beneficence, and justice. As Freedman and Shapiro observe, the principles provide a potentially comprehensive way of dividing up moral questions related to research.[44] Respect for persons can be viewed as covering issues related to the *rights of research subjects*. Beneficence embraces concerns related to the *welfare*

of research subjects. Finally, justice may be viewed as a catch-all category, particularly of issues with *broad social implications*.

In practice, however, research ethicists have focused on a narrow range of issues in clinical research. In large part, this phenomenon is historically rooted. I have argued elsewhere that ethical interest and effort was largely shaped by a succession of research scandals in Europe and North America.[45] Reacting to the particular issue (or issues) highlighted by each of these scandals, ethicists focused their efforts on only a few areas, e.g., informed consent, confidentiality, assessment of risk-benefit. This selective interest is evident from the literature. Freedman and Shapiro review articles cited in the cumulative index (1979-1990) of the only peer-review journal devoted to research ethics, *IRB: A Review of Human Subjects Research*.[46] The largest number of articles relate to news items and federal regulations (n=100); articles on informed consent (99), confidentiality (43) and risk-benefit assessment (40) make up the next largest groups. Only a very few deal with research design in general (20) and subject selection procedures in particular (5).

The analysis of ethical issues in research need not be at the mercy of the scandal *du jour*. An alternative is a *comprehensive ethical analysis* of issues in clinical research. Freedman and Shapiro explain:

Principles crafted in reaction to scandal respond to the question:

> How can this evil be avoided? What is needed instead is to ask, How can our conduct of research be more ethically sensitive, be, in fact, improved? This alternative approach must begin by acknowledging that each important choice taken in the design and conduct of human research is of potential ethical interest. Ethics, as practical philosophy, deals with the evaluation of human choice: What option is required by moral duty; what action should be taken? And research with human subjects is replete with clinical, biological, and statistical choices: What will be tested? How will it be tested? Who will be tested? And how will they be recruited? When will the test be complete? Each of these choices is as susceptible to ethical reflection and critique as is any aspect of a clinical trial. From the point of view of ethics as the evaluation of choice, there is no inherent distinction between some aspects of trials that raise ethical questions (like the consent form or payments to subjects or investigators) and ethically neutral elements.[47]

This thesis undertakes a part of this comprehensive approach to the ethical analysis of clinical research. Not motivated by scandal (or fashion) we undertake an examination of a little-studied part of the clinical trial, criteria for trial eligibility. We do so motivated by the believe that *all parts of the clinical trial*

protocol present ethical issues (in so far as they represent the end product of human choice). We ask: How can the conduct of clinical research be improved?

We could have focused on any one of the above questions, but we chose to restrict our enquiry to the query, "Who will be tested?" It is surely a question with scientific implications: explanatory trialists argue that good science demands a homogeneous patient population, trial pragmatists argue -- more convincingly, I think -- that a heterogeneous patient population is, for a variety of reasons, preferred scientifically. But, it is also a question with ethical implications. In chapter 1 we argued that clinical research should not prey upon the vulnerable, nor should it exclude without good reason those who may benefit from research participation. Recently, the applicability of the results of research findings has recently received considerable ethical (and political) attention. In chapter 5, I presented a comprehensive discussion of Walzer's political philosophy and the just allocation of the knowledge arising from clinical research.

While each of the above ethical issues is important, the question, "Who will be tested?" and what difference that makes ethically, gets at the fundamental ethical justification for clinical research. Under what circumstances may a randomized trial be properly initiated? And most important to us here, what is the purpose of such a randomized clinical trial? What are the implications of this for the selection of subjects for clinical research? Freedman's concept of clinical

equipoise addresses all of these questions.

When can a trial be legitimately initiated? There is a consensus that at the beginning of a randomized clinical trial comparing two (or more) treatments an honest null hypothesis must exist.[48] In other words, uncertainty must exist as to the relative merits of the treatments being tested in the trial.[49] Some have argued that this means that the evidence on behalf of the two treatments must be *precisely balanced* — a notion referred to as 'theoretical equipoise.'[50] Freedman has correctly pointed out that this understanding of equipoise is too fragile to be of any practical value:

> Theoretical equipoise is overwhelmingly fragile; that is, it is disturbed by a slight accretion of evidence favoring one arm of the trial...[It] is also highly sensitive to the vagaries of the investigator's attention and perceptions. Because of its fragility, theoretical equipoise is disturbed as soon as the investigator perceives a difference between the alternatives — whether or not any genuine difference exists...Finally...[it] is personal and idiosyncratic. It is disturbed when the clinician has...what 'might even be labeled a bias or a hunch,' a preference of a 'merely intuitive nature.'"[51]

Freedman argues persuasively for different understanding of equipoise termed 'clinical equipoise.'

In order for a trial to be initiated ethically, a state of clinical equipoise must exist at the trial's inception; that is, there must be honest, professional disagreement among expert medical practitioners as to the preferred treatment.[52] This disagreement can arise in a number of different ways. For example, two different yet standard treatments may be advocated by separate groups of expert practitioners. Members of each group may well recognize that there is evidence to support both treatments; however, they find the studies supporting their treatment of choice to be the most convincing. Another example: if a single standard treatment exists, evidence may come to light that a new treatment may be preferable, for example, it may be more effective, or equally effective but associated with less side effects.

Whatever the cause of the state of clinical equipoise, a trial is initiated to resolve the real (or potential, in the case of a very new treatment) disagreement among expert medical practitioners. In short, the purpose of a clinical trial is to change medical practice. The second ethical prerequisite for a randomized clinical trial is that

> the trial must be designed in such a way as to make it reasonable to expect that, if it is successfully concluded, clinical equipoise will be disturbed. In other words, the results of a successful trial should be convincing enough to resolve the dispute among clinicians.[53]

The impetus for the trial arises out of uncertainty as the preferred treatment; the trial is conducted to decide which of the treatments is to be preferred in clinical practice.

Given that clinical equipoise supports the notion of the randomized clinical trial as an instrument to change clinical practice, equipoise supports a pragmatic approach to trial design. Recall from our discussion of explanatory and pragmatic trial designs that pragmatic trials, studies which include heterogenous groups of patients are designed to "compare two treatments under the conditions in which they would be applied in practice...[they seek] to answer the question — which of the two treatments should we prefer?"[54] Explanatory trials, trials with narrow selection criteria and homogeneous study populations, bear little direct relevance to clinical reality. Freedman:

> This 'fastidious' [explanatory] approach purchases scientific manageability at the expense of an inability to apply the results to the 'messy' conditions of clinical practice... Overly 'fastidious' [explanatory] trials, designed to resolve some theoretical question, fail to satisfy the second ethical requirement of clinical research, since the special conditions of the trial will render it useless for influencing clinical decisions, even if it is successfully completed."[55]

Clinical equipoise offers us an explanation of the preconditions for an

ethical clinical trial — disagreement among expert practitioners — and the purpose of a clinical trial — to resolve that disagreement; to change clinical practice. If a clinical trial is to accomplish this task optimally the patients included in that study must be reasonably representative of patients in clinical practice. Thus, a pragmatic approach to the selection of subjects for clinical research is preferred.

The applicability of the results of clinical trials to clinical practice is a point of overlap between scientific and ethical issues in clinical trials. Clinical trial pragmatists object to highly restrictive eligibility criteria on the basis that "[i]f the patients in a clinical trial are not representative of the entire patient population...the generalizability of the results to the entire patient population may be compromised."[56] The trial's impact on clinical practice may be harmed by an unduly homogeneous study population and, certainly, information on the proper treatment of important patient subgroups will not be forthcoming. Equipoise requires that a trial, if successful, resolve disagreement among expert practitioners — change practice. Thus, the ethicist has *precisely* the same concerns. *Clinical scientists and ethicists are united by the vision of the phase III clinical trial as an instrument of changing medical practice.*

Methodology used in this thesis and suggestions for further research

Clinical equipoise tells us not only when it is ethical to initiate a clinical trial it also tell us what the purpose of that trial ought to be — to resolve disagreement among expert practitioners; to change clinical practice. Thus, equipoise provides us with a teleology of the randomized clinical trial; it tells us the end — τα προς τα τελος — that an ethically and scientifically sound trial ought to achieve. An agricultural analogy: as the acorn is to the oak tree, so too the clinical trial to its end, altering clinical practice. The implications of this are both profound and far reaching for the design, conduct and reporting of clinical trials. We said before that a comprehensive approach to the ethical analysis of clinical research sees all aspects of the planning, conduct and reporting of research as having ethical elements (in so far as human choice is involved in those aspects of trials). Clinical equipoise provides us with a tool to *systematically examine the ethics of clinical research.*

Clinical equipoise was the lens through which we viewed the problems presented in this thesis. When asking about changes in eligibility criteria in clinical trial protocols, equipoise caused us to focus on the implication for those changes on the generalizability of trial results. When examining the interpretation of criteria by investigators, equipoise directed our attention to the ramifications for the interpretation of trial results when ambiguous criteria were used to pick the

study population. Finally, when studying the reporting of criteria, it was equipoise that lead us to be critical of trials that failed to fully disclose criteria (how can clinicians properly assess the impact of the results on their own practice?) or contain unnecessary criteria (the applicability of study results may be restricted unduly). Thus, the approach in this thesis is an application of the principle of clinical equipoise to one aspect of clinical research, namely, the selection of subjects for research participation. Methodologically it is, therefore, *systematic,* insofar as it analyses a range of ethical problems in accord with a single principle, clinical equipoise, and a part of a larger *comprehensive* approach that sees the need for ethical analysis in all aspects of the conduct of clinical research.

The analysis in chapters two through four is characterized by a combined theoretical and empirical approach to ethical analysis. Our theoretical understanding of the problem, informed by clinical equipoise, gives us a sense of how things ought to be; an empirical examination informs us of how they are; the empirical facts then feedback to inform the theoretical understanding. Moral theory informs our analysis of practice; and the observed phenomena inform our theoretical understanding. For example, in the second chapter, *a priori* clinical equipoise tells us that criteria for trial eligibility ought to be minimized. But how far may criteria be minimized and still allow for meaningful interpretation of trial results? Examining the phenomena, we were surprised by the fact that the POG

trials contained no safety criteria whatsoever. Given the high quality of the POG trials and the important influence that these trial have had on clinical practice, we conclude that such criteria are not necessary for high quality research. Theory informs practice; practice informs theory.

Given this understanding of the methodology utilized in this thesis, a wide variety of questions for further practice are suggested. I will list but a few.

● In chapters two through four subject selection procedures were examined at three stages in the genesis and dissemination of medical research: clinical trial protocol, interpretation by investigators, and reporting in study communications. Selection procedures ought to be studied at yet another important point: the interpretation by clinicians in practice. As we have seen, many clinical trials in oncology use large numbers of restrictive criteria. How do physicians in practice interpret the results of these trials? Do they only apply the study results (i.e., the treatment in question) to patients in their practice who would have been eligible for the trial? Or, as we suspect, do they disregard some of the trial's eligibility criteria in their inference? What kinds of criteria do they ignore? What are the implications for the design of clinical trials? This study would likely involve a questionnaire using clinical vignettes (not unlike that used in chapter 3) to oncologists in clinical practice. Clinicians could be presented with trial results (either very recent or hypothetical) and asked if they would apply the treatment in various clinical situations

constructed around representative eligibility criteria from each of the categories in the schema (chapter 2).

● Another important question stems from our study of the reporting of eligibility criteria in study communications. As we indicated in chapter 4, previous studies of the reporting of study methods in study communications have focused on the *clarity* of reporting in journal articles. Our innovation was to look at the *accuracy* of reporting of methods by comparing the information found within the clinical trial protocol with that found in the journal article. Furthermore, we recognize that the journal article is not the sole medium for the reporting of clinical trial methods. Accordingly we examine methods papers and Clinical Alerts issued by the U.S. National Institutes of Health. The study presented in chapter four could be broadened to look at the reporting of other aspects of trial methods, including study hypotheses, study power, minimally important clinical effect, study interventions and analytic techniques. If we were to document a similar loss of information as documented in the study on the reporting of eligibility criteria, important consequences would follow for the reporting of study results.

Clinical equipoise -- as the instrument for the systematic analysis of the ethics of clinical research — has broad implications for the conduct of research. Two of these areas — many other could be suggested here — for further

examination are as follows:

● Framing of study hypotheses. Traditionally, hypotheses in clinical research are framed around a null hypothesis of zero difference. A given treatment is concluded to be superior to a comparison treatment if the lower bound of the 95% confidence interval for the difference between the two treatments excludes zero. Thus, typically, a new treatment is said to be superior to another if there is strong evidence (i.e., statistical significance) of a non-zero difference between the two treatments. In some cases, evidence of a non-zero difference between treatments may indeed be sufficient to change clinical practice.

In other cases, particularly when the experimental treatment is toxic, complicated to administer or expensive, evidence of a non-zero difference between treatments is unlikely to sway the practice of expert clinicians. In these cases clinical trials must take the difference necessary to sway practice into account in the study hypotheses. A successful clinical trial then will provide strong evidence to rule in or rule out a treatment difference greater than this minimal value. Thus, the null hypothesis of no-treatment difference ought to be replaced with a null hypothesis that the treatment difference is equal to or less than minimal difference to shift clinical practice. How to define this minimal difference? How will this difference affect the size and practicality of clinical trials? Clinical equipoise, as defined above, seems to only allow for the inclusion of the beliefs of physicians, is

there an appropriate role for patients and patient communities in these determinations?

● Interim analysis. Clinical trials that involve death or serious morbidity as the primary outcome often plan for one or more interim analyses prior to the planned conclusion of the study. If the observed difference between the treatments positive (or negative) is unexpectedly large, an interim analysis allows trial investigators to stop the trial early so that subjects are not unnecessarily exposed to inferior treatment (be it the experimental or the control treatment). A variety of statistical methods have been developed to (1) ensure that the interim analysis does not alter the type I error rate for the study and (2) to ensure that more conservative stopping rules are employed earlier in the study to protect from prematurely stopping the trial. Despite the importance of these contributions, statistical 'stopping rules' fail to capture all of the concerns that ought to enter into the decision to terminate a trial prematurely. Other factors, including the degree of skepticism of clinicians to the trial result, the quality of the data, the amount of data in the "pipeline" (collected but not included in the analysis), must be considered. How might clinical equipoise help to shape this decision making process? What impact will non-zero null hypotheses have on statistical 'stopping rules'?

Given these few examples of further projects stemming from the work presented in this thesis, it is clear that the student of the systematic and comprehensive ethical

analysis of clinical research has much with which to occupy herself.

Claims to originality

Each of the four experimental chapters (chapters two through five) contains elements which represent an original contribution to the literature. The original contributions are as follows:

- The examination and characterization of changes in eligibility criteria in two diachronous samples of cancer clinical trials (chapter 2).
- The development and validation of a schema to classify eligibility criteria (chapter 2).
- The survey of the interpretation of eligibility criteria by clinical trial investigators (chapter 3).
- The development and validation of the study instrument for this study (chapter 3).
- The examination of the *accuracy* of the reporting of eligibility criteria in methods papers, journal articles, and Clinical Alert issued by the U.S. National Institutes of Health (chapter 4) and the characterization of information loss with the schema developed in chapter 2.
- The application of Walzer's political philosophy to define the just distribution of knowledge resulting from clinical research, including the classification of unjust

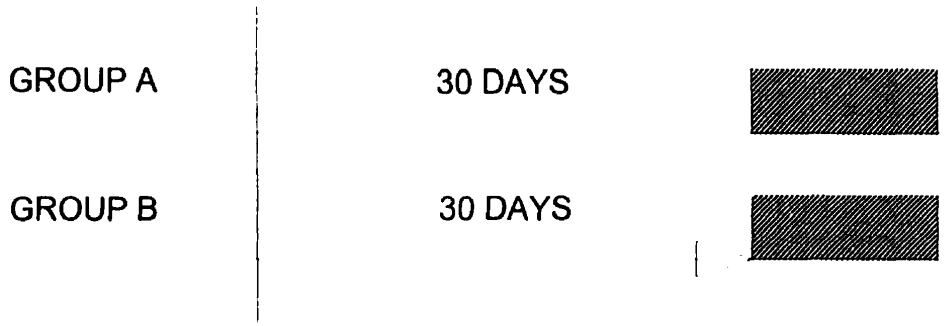
inclusion in and exclusion from research as two instances of the same phenomenon -- domination (chapter 5).

Conclusion

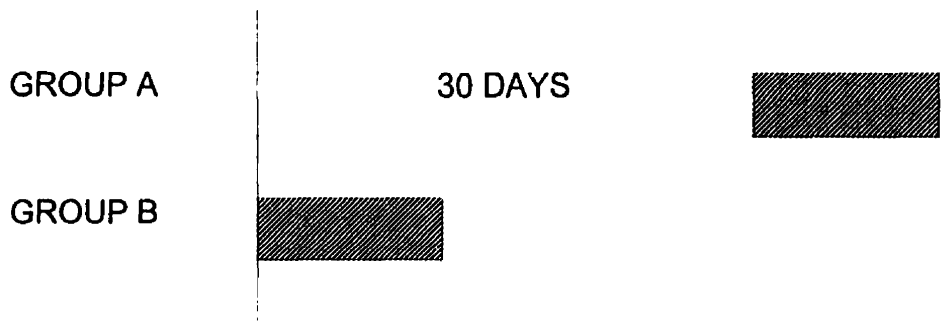
In this thesis we have addressed only one aspect of the range of questions suggested by a comprehensive approach to ethical analysis of clinical research. We limited our scope to the question, "Who will be tested?" In chapter 1 we provided the reader with a historical overview of the ethical issues seen to be of importance in the selection of subjects for clinical research. In chapter 2 we examined change in eligibility criteria over time in two important sets of protocols. In chapter 3 we examined the interpretation of study criteria by clinical investigators. In chapter 4 we looked at the accuracy of reporting of criteria in study communications. In chapter 5 we tried to provide one possible comprehensive philosophical account of just selection procedures for research participation. Finally, in the last chapter, we have seen that the question, "Who will be tested?," has both scientific and ethical components. Components that at points seem inextricably entangled.

The most important aspect of this thesis is, I believe, that it is, self-consciously, an instantiation of a comprehensive and systematic approach to the ethical analysis of clinical research. Implicitly, it is a rejection of the notion that the legitimate domain of ethical analysis of research is restricted only to certain

aspects of the research protocol (e.g., the consent form). It is also a rejection of the idea that ethical analysis of research ought to be limited to the research protocol itself and not extend to the conduct and dissemination of research. Finally, it is rejection of the notion that ethical analysis is wholly separate from scientific methodology. We have mapped out but one area of overlapping concern between science and ethics -- subject selection and the applicability of research results. *In toto*, these three realizations open a world of possibility for ethical analysis in clinical research. The work presented here is, therefore, but the beginning of a beginning.



EXPLANATORY DESIGN



PRAGMATIC DESIGN

- RADIOTHERAPY

- DRUG UNDER STUDY

Figure 1. Explanatory and pragmatic clinical trial designs. In the explanatory design, treatment with a radiosensitizing drug followed by radiotherapy is compared with a thirty day no-treatment period followed by radiotherapy (i.e., delayed radiotherapy). In the pragmatic design, treatment with a radiosensitizing drug followed by radiotherapy is compared with immediate radiotherapy. (Figure taken from Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. *Journal of Chronic Disease* 1967; 20: 637-648.)

Arguments for explanatory selection procedures	Arguments for pragmatic selection procedures
● Including patients of differing prognosis (i.e., a heterogeneous study population) will increase variability in a study and, hence, decrease its power.	● A truly homogeneous patient population cannot be defined; patient-to-patient variability is the largest source of variation.
● Studies involving homogeneous groups of patients are more efficient (in terms of "residual sample size requirements").	● A study's power is maximized by removing eligibility criteria and making the sample size larger, not further restricting the study population.
● Within a heterogeneous study population, qualitative differences among subgroups may cancel one another out, thus obscuring the treatment's "true" effect.	● Trials with highly restrictive eligibility criteria may have difficulty accruing adequate numbers of patients.
● In a heterogeneous study population, there are insufficient numbers in individual subgroups to do a meaningful statistical analysis.	● Added eligibility criteria increase trial complexity and costs; minimizing criteria reduce complexity and cost.
● Studies with broad inclusion criteria (and, hence, numerous subgroups of potential interest) may lead to misleading multiple subgroup analyses.	● Studies involving heterogeneous groups of patients are more efficient (in terms of accrual rate and time to completion).
	● Qualitative differences among subgroups are, generally speaking, uncommon.
	● Statistical techniques exist to adjust for important known covariates.
	● The results of narrowly-focused studies may not be applicable to the patient population at large; trials with heterogeneous patient populations are more likely to be widely applicable in clinical practice.

Table 1. Summary of arguments for explanatory selection procedures and reasons for pragmatic selection procedures for clinical trials. (See text for a detailed explanation of each point.)

Table 2.

Treatment/ disease setting	Outcome measure		
	Death	Non-fatal infarction	Non-fatal stroke
Antiplatelet Therapy			
1. acute MI	-23% ± 4*	-49% ± 9	-46% ± 17
2. long-term following MI	-15 ± 5*	-31% ± 5	-42% ± 11
3. cerebrovascular disease	-15% ± 7*	-35% ± 12	-22% ± 7
4. unstable angina	-42% ± 15	-40% ± 15	incomplete data
Beta-blockers			Cardiac arrest/ sudden death
1. acute MI	-14% ± 6	-18% ± 7	-15% ± 7
2. long-term following MI	-22% ± 4	-27% ± 5	-32% ± 5
3. threatened MI	inadequate data	-13% ± 6	inadequate data
Calcium Blockers			
1. acute MI	+10% ± 11	not available	not available
2. long term following MI	+6% ± 10	-9% ± 7	not available
3. unstable angina	+70% ± 50	0% ± 12	not available

* - vascular deaths; MI - myocardial infarction.

Table 2. Reductions (% ± standard error) observed with various treatments for vascular disease, demonstrating similarity of the direction of effect on the same endpoint in quite different populations. (Table taken from Yusuf S, Held P, Teo KK. Selection of patients for randomized controlled trials: implications of wide or narrow eligibility criteria. *Statistics in Medicine* 1990; 9: 73-86.)

References

1.Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. *Journal of Chronic Disease* 1967; 20: 637-648.

2.*ibid.*.

3.*ibid.*.

4.*ibid.*.

5.*ibid.*.

6.Feinstein AR. An additional basic science for clinical medicine: II. The limitations of randomized trials. *Annals of Internal Medicine* 1983; 99: 544-550.

7.*ibid.*.

8.Weijer C. Characterizing the population in clinical trials: barriers, comparability, and implications for review. Master's thesis. McGill University, Montreal, 1995.

9.Anderson KW, Mouridsen HT. Danish Breast Cancer Cooperative Group (DBCG): a description of the register of the nation-wide programme for primary breast cancer. *Acta Oncologica* 1988; 27: 627-647.

10.Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient (parts I and II). *British Journal of Cancer* 1976; 34: 585-612 and 1977; 35: 1-39.

11.Simon R. Design and conduct of clinical trials. In: DeVita VT, Hellman S, Rosenberg S (eds.). *Cancer: Principles and Practice of Oncology*. Philadelphia:

J.B. Lippincott Company, 1989: 396-420.

12.Payer L. *Medicine and Culture: Varieties of Treatment in the United States, England, West Germany, and France*. New York: Penguin Books, 1989.

13.Hill AB. The clinical trial. *British Medical Bulletin* 1951; 7: 278-282.

14.Gail MH. Eligibility exclusions, losses to follow-up, removal of randomized patients, and uncounted events in cancer clinical trials. *Cancer Treatment Reports* 1985; 69: 1107-1113.

15.Sylvester R. Planning cancer clinical trials. In: Buyse ME, Staquet MJ, Sylvester RJ (eds.). *Cancer Clinical Trials: Methods and Practice*. New York: Oxford University Press, 1984: pages 47-63.

16.Sackett DL. On some prerequisites for a successful clinical trial. In: Shapiro SH, Louis TA. *Clinical Trials: Issues and Approaches*. New York: Marcel Dekker, Inc., 1983: pp. 65-79.

17.Simon R. Patient heterogeneity in clinical trials. *Cancer Treatment Reports* 1980; 64: 405-410.

18.Simon R. Design and conduct of clinical trials. In: DeVita VT, Hellman S, Rosenberg S (eds.). *Cancer: Principles and Practice of Oncology*. Philadelphia: J.B. Lippincott Company, 1989: 396-420.

19.Sylvester R. Planning cancer clinical trials. In: Buyse ME, Staquet MJ, Sylvester RJ (eds.). *Cancer Clinical Trials: Methods and Practice*. New York: Oxford University Press, 1984: pages 47-63.

20.Simon, *op. cit.*, note 18.

21.Yusuf S, Held P, Teo KK. Selection of patients for randomized controlled trials: implications of wide or narrow eligibility criteria. *Statistics in Medicine* 1990; 9: 73-86.

22.Begg CB, Engstrom PF. Eligibility and extrapolation in cancer clinical trials. *Journal of Clinical Oncology* 1987; 5: 962-968.

23.Yusuf 1990, *op. cit.*.

24.Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Statistics in Medicine* 1984; 3: 409-420.

25.Peto R. Clinical trial methodology. *Biomedicine* 1978; 28: 24-36.

26.Yusuf 1990, *op. cit.*.

27.Yusuf 1990, *op. cit.*.

28.Peto R, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. *Annals of the New York Academy of Sciences* 1993; 703: 314-340.

29.Peto 1993, *op. cit.*.

30.Peto 1993, *op. cit.*.

31.George SL. Reducing patient eligibility criteria in cancer clinical trials. *Journal of Clinical Oncology* 1996; 14: 1364-1370.

32.George, *op. cit.*.

33.Yusuf 1990, *op. cit.*.

34.Buyse ME. The case for loose inclusion criteria in clinical trials. *Acta Chir Belg* 1990; 90: 129-131.

35.George, *op. cit.*.

36.Yusuf 1990, *op. cit.*.

37.Peto 1993, *op. cit.*.

38.Begg and Engstrom, *op. cit.*.

39.Leventhal BG, Wittes RE. *Research Methods in Clinical Oncology*. New York: Raven Press, 1988.

40.Armitage P. Exclusions, losses to follow-up, and withdrawals in clinical trials. In: Shapiro SH, Louis TA. *Clinical Trials: Issues and Approaches*. New York: Marcel Dekker, Inc., 1983: pp. 99-113.

41.George, *op. cit.*.

42.Stenning S. 'The uncertainty principle': selection of patients for cancer clinical trials. In: Williams CJ. *Introducing New Treatments for Cancer: Practical, Ethical and Legal Problems*. New York: John Wiley & Sons, 1992: pp. 161-172.

43.Stenning, *op. cit.*.

44.Freedman B, Shapiro S. Ethics and statistics in clinical research: towards a more comprehensive examination. *Journal of Statistical Planning and Inference* 1994; 42: 223-240.

45.Weijer C. Evolving ethical issues in the selection of subjects for clinical research. *Cambridge Quarterly of Healthcare Ethics* 1996; 5: 334-345.

46.Freedman and Shapiro, *op. cit.*.

47.Freedman and Shapiro, *op. cit.*.

48. Levine RJ. *Ethics and Regulation of Clinical Research*. New Haven: Yale University Press, 1988: pp. 187-190.

49. Fried C. *Medical Experimentation: Personal Integrity and Social Policy*. Amsterdam: North Holland Publishing, 1974.

50. Shaw LW, Chalmers TC. Ethics in cooperative clinical trials. *Annals of the New York Academy of Sciences* 1970; 169: 487-495.

51. Freedman B. Equipoise and the ethics of clinical research. *New England Journal of Medicine* 1987; 317: 141-145.

52.Freedman, *op. cit.*.

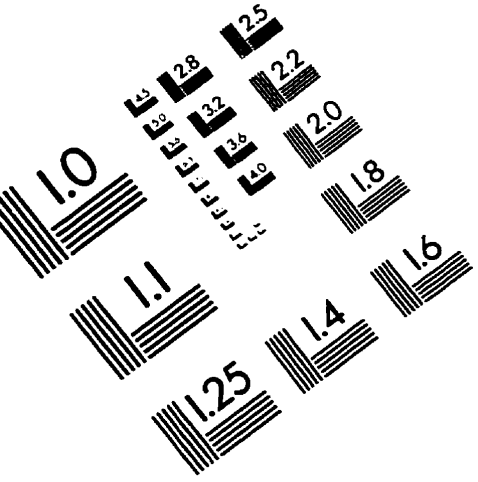
53.Freedman, *op. cit.*.

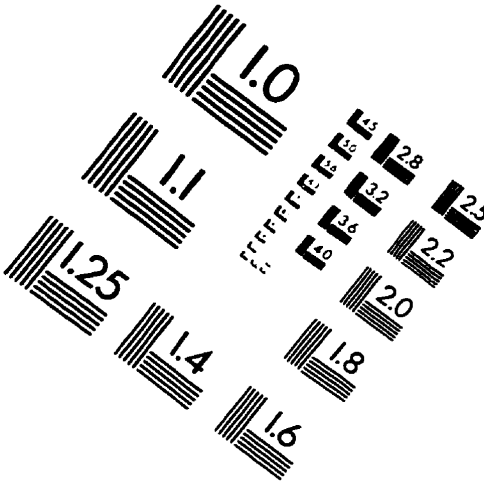
54.Schwarz and Lellouch, *op. cit.*.

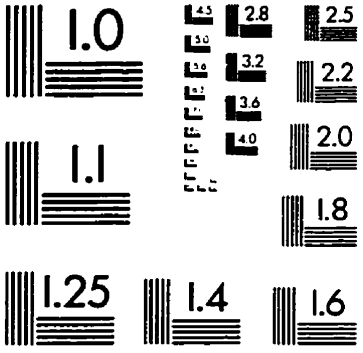
55.Freedman, *op. cit.*.

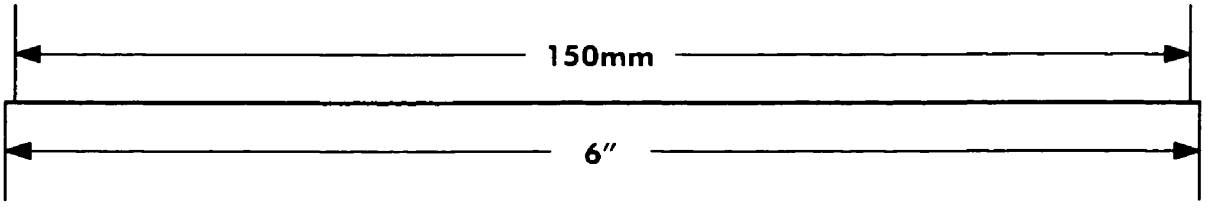
56.Leventhal, *op. cit.*.

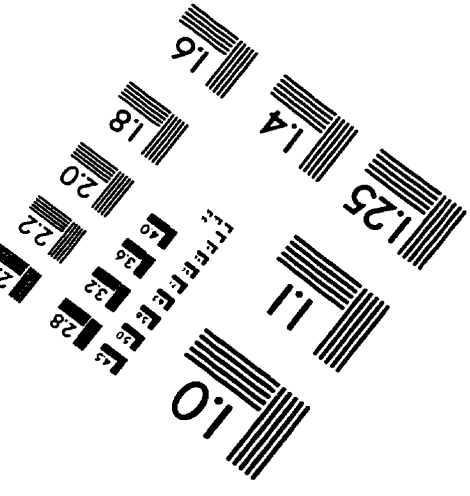
IMAGE EVALUATION TEST TARGET (QA-3)

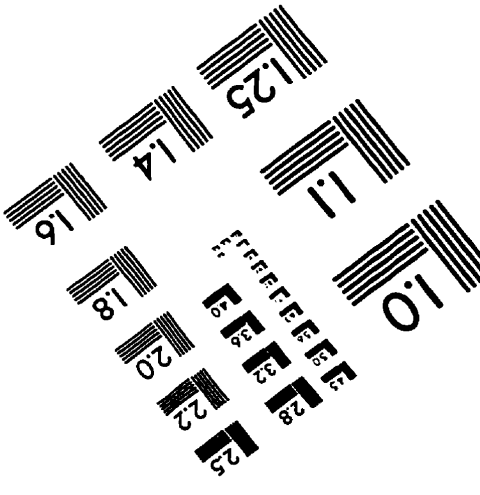












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989