7-1995

Characterizing the Population in Clinical Trials: Barriers, Comparability, and Implications for Review

Charles Weijer
McGill University, cweijer@uwo.ca

National Library of Canada

Acquisitions and Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada

Direction des acquisitions et des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Canada

CHARACTERIZING THE POPULATION IN CLINICAL TRIALS: BARRIERS, COMPARABILITY, AND IMPLICATIONS FOR REVIEW

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

July 1995

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Master of Science in Experimental Medicine, Specialization in Bioethics.

ISBN 0-612-08058-7

Canada

Abstract (English)

The definition of the study population for a clinical trial via the criteria for trial eligibility has implications for the validity of the study and its applicability to clinical practice. Though issues of equity regarding the selection of subjects for research have long been a concern of ethicists, issues regarding the impact of subject selection on a trial's generalizability have only recently attracted ethical scrutiny. After a review of the history of the ethics of subject selection, I focus on three empirical questions regarding the generalizability of clinical trials. (1) What proportion of diseased populations are studied in clinical trials? (2) How are subjects selected for clinical trial participation (and what are the main barriers to participation)? (3) Are clinical trial participants comparable to non-participants? Finally, the role of the Institutional Review Board -- Research Ethics Board in Canada -- in assessing the generalizability of clinical research is discussed.

Abstract (French)

Définir la population étudiée dans le cadre d'un essai clinique par le truchement des critères d'admissibilité à l'essai revêt plusieurs conséquences sur la validité de l'étude et ses possibilités d'application clinique. Le problème de l'équité dans le choix des sujets aptes à la recherche préoccupe depuis longtemps les éthiciens mais l'intérêt qu'ils prêtent aux questions liées à l'impact de la sélection des sujets sur le potentiel généralisable des essais est récent. Après un passage en revue de l'évolution de l'éthique de la sélection des sujets, je m'intéresse à trois questions empiriques liées au potentiel généralisable des essais cliniques : (1) Quelle est la proportion de populations malades étudiées dans les essais cliniques? (2) Comment les sujets sont-ils sélectionnés pour les essais cliniques (et quels sont les principaux obstacles à leur participation)? (3) Les participants aux essais cliniques se comparent-ils aux non-participants? Enfin, j'examine en détail le rôle des comités d'éthique de la recherche dans l'évaluation du caractère généralisable de la recherche clinique.

Preface

In accordance with the Guidelines Concerning Thesis Preparation I have taken the option, according to section (7), of writing the experimental part of the thesis (Chapters two and three) in the form of original papers submitted for publication to learned journals. This provision reads as follows:

> The candidate has the option, subject to the approval of the Department, of including as part of the thesis the text, or duplicated text (see below), of an original paper, or papers. In this case the thesis must still conform to all other requirements explained in the Guidelines Concerning Thesis Preparation. Additional material (procedural and design data as well as descriptions of equipment) must be provided in sufficient detail (e.g., in appendices) to allow a clear and precise judgement to be made of the importance and originality of the research reported. The thesis should be more than a mere collection of manuscripts published or to be published. It must include a general abstract, a full introduction and literature review and a final overall conclusion. Connecting text which provide logical bridges between different manuscripts are usually desirable in the interests of cohesion.
>
> It is acceptable for theses to include as chapters authentic copies of papers already published, provided these are duplicated clearly on regulation thesis stationary and bound as an integral part of the thesis. Photographs or other materials which do not duplicate well must be included in their original form. In such instances, connecting texts are mandatory and supplementary explanatory material is almost always necessary.
>
> The inclusion of manuscripts co-authored by the candidate and others is acceptable but the

> candidate is required to make an explicit statement on who contributed to such work and to what extent, and supervisors must attest to the accuracy of such claims, e.g. before the Oral Committee. Since the task of the Examiners is made more difficult in these cases, it is in the candidate's interest to make the responsibilities of authors perfectly clear. Candidates following this option must inform the Department before it submits the thesis for review.

Thus, each chapter of this thesis bears its own Abstract, Introduction, Methods, Results, Discussion and References. Also, as required by the Guidelines, there is a common abstract, a general introduction (Chapter one) and a general discussion (Chapter four) as well as claims to originality and suggestions for further research.

The submitted manuscripts are as follows:

- Chapter 2. Weijer C. Eligibility criteria and other barriers to enrollment in randomized controlled trials. (Submitted for publication).
- Chapter 3. Weijer C. Are randomized controlled trial participants comparable to non-participants? -- A review of the empirical literature. (Submitted for publication).

The candidate was responsible for all of the work in both of these papers. Helpful advice and comments on earlier versions of each of the 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., Trudo Lemmens, LL.L. and Myriam Skrutkowska, B.Sc.N..

The candidate's work is supported by a fellowship from the Medical Research Council of Canada. The candidate would like to express his sincere gratitude to his colleagues in the Clinical Trials Research Group for their encouragement and support during the preparation of this thesis. He would also like to thank Anthony Belardo, B.A. and Elena Plotkin who both provided editorial assistance with the thesis manuscript. The candidate is particularly grateful to his supervisors, Professors Benjamin Freedman and Abraham Fuks, for their guidance, both intellectual and moral.

Wisdom has built her mansion,
and set up her seven pillars;
her beasts are slain, her wines are blended,
her table is prepared;
she has sent her maidens out to cry,
on the thoroughfares of the city,
"Let all who are heedless turn in here!"
She calls to him who is devoid of sense,
"Come, eat my bread,
drink wines that I have blended;
leave your foolish ways and live,
follow the ways of thoughtful sense."
[Proverbs 9:1-6]

Table of Contents

Chapter 1:

Introduction: Setting the stage

Introduction

Issues involving the selection of subjects for clinical research have occupied clinical trial designers and ethicists for at least the last three decades. The selection of subjects for clinical research is defined by criteria for eligibility in the research protocol. When defining the eligible population for a study, clinical trial designers attempt to strike a balance between defining a homogeneous study population (enhancing the validity of the study) and including a diverse enough group of subjects such that the results will be applicable to clinical practice (enhancing the study's generalizability). Until relatively recently, ethicists and research ethics committees -- Research Ethics Boards (REBs) in Canada; Institutional Review Boards (IRBs) in the United States -- had concerns regarding procedures for the selection of subjects that, by and large, did not overlap with those of trial designers.

Until recently, ethicists were largely concerned with issues of justice in the selection of subjects: the equitable distribution of the burdens and benefits of research (discussed in detail below). The impact of the criteria for research eligibility on the generalizability of the research findings was not considered an ethical issue. Recent regulatory changes, including the NIH Guidelines on

the Inclusion of Women and Minorities as Subjects in Clinical Research (1994), have given IRBs a mandate to evaluate the eligibility criteria of research proposals with regard to their impact on the generalizability of the research findings.[1] These regulatory changes have been paralleled by efforts to systematize approaches to the ethical analysis of protocols which call for the scrutiny of eligibility criteria on the basis of their impact on generalizability.[2]

I have therefore undertaken to study the impact of eligibility criteria on the generalizability of randomized controlled trials. Like any new intellectual tack, the questions proposed and the results obtained must be viewed in context. The context here is at least two fold: normative and historical. The predominant normative context for research ethics is set out by the Belmont Report.[3] The historical context encompasses the shifting pattern of ethical concerns relating to the selection of subjects for research alluded to above. The purpose of this chapter is to review these two subject areas, setting the stage for what is to follow. The last brief section of this chapter will list the study questions addressed in this thesis.

The Belmont principles

In the Belmont Report, the members of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (hereafter, "National Commission") set out three ethical principles that could serve to describe the foundation of research ethics.[3] These three principles -- respect for persons, beneficence, and justice -- are "generally accepted in our cultural tradition" and were chosen for their particular relevance to the ethics of research.[3] Understanding that rules regarding the ethical conduct of human experimentation would, at times, conflict with each other, the National Commission intended that the principles would "provide an analytical framework that will guide the resolution of ethical problems arising from research involving human subjects".[3]

Respect for persons. The principle of respect for persons entails two ethical requirements. First, that the choices of an autonomous person -- "an individual capable of deliberation about personal goals and of acting under the direction of such deliberation" -- ought not to be interfered with unless those choices are likely to harm others.[3] Second, that persons who are not capable of autonomous choice, including children, the mentally ill, and (perhaps) persons who are incarcerated, are entitled to

special protection.[3] In the context of research, the principle of respect for persons is taken to require that subjects of research must consent to participate. Consent for research participation must be voluntary (i.e., not coerced), informed (the subject must be informed of the purpose of the research, the procedures involved, alternatives, etc.), and comprehending (efforts must be undertaken to ensure that the subject understands the information relayed).

Beneficence. The principle of beneficence in general requires that efforts must be undertaken both to protect persons from harm as well as to attempt to ensure their wellbeing. The Belmont Report defines two complementary rules that define beneficent actions: beneficent actions "(1) do no harm and (2) maximize possible benefits and minimize possible harms".[3] Thus, in the context of research, the investigators and IRBs must assess the risks and benefits that research presents to potential subjects to ensure that a favourable risk-benefit ratio exists. The principle also requires that potential research benefits to society be considered. "Beneficence thus requires that we protect against risk of harm to subjects and also that we be concerned about the loss of substantial benefits that might be gained from research".[3]

Justice. The principle of justice demands that burdens and benefits be distributed fairly. Conceptually, goods can be justly distributed on the basis of a number of different formulations such as "to each person an equal share" or, "to each person according to individual need" and so on. (For a detailed discussion of the different formulations of distributive justice see Tom Beauchamp's chapter in the appendix to the Belmont Report).[4] In the context of research ethics, justice demands that the burdens and benefits of research be distributed equitably.

IRBs must scrutinize the procedures for the selection of subjects for research to ensure justice on two levels: individual and societal. On the level of the individual, researchers "should not offer potentially beneficial research only to some patients who are in their favor or select only 'undesirable' persons for risky research".[3] On a societal level, justice requires that classes of persons who are already burdened in some way ought not be further burdened by research unless it is necessary to do so.

> Thus, it can be considered a matter of social justice that there is an order of preference in the selection of subjects (e.g., adults before children) and that some classes of potential subjects (e.g., the institutionalized mentally infirm or prisoners) may be involved as research subjects, if at all, only on certain conditions.[3]

Evolving ethical concerns in the selection of subjects for clinical trials.

Wittgenstein, in his famous critique of logical positivism, noted that the influence of an idea can be such that it alters the way that we see the world. "It is like a pair of glasses on our nose through which we see whatever we look at", he said. "It never occurs to us to take them off".[5] This view of the power of an idea suggests that the interpretation of an event, and what response this event calls for, can depend upon the view one has of the world -- if you will, upon the pair of glasses through which you are viewing the event. A person who is naive about medical facts may, for example, interpret chest pain upon exertion as be a sign that he is "over doing it"; were he more medically knowledgeable, the same symptom might be interpreted as a possible indicator of coronary artery disease. The naive interpretation calls for rest; the informed interpretation calls for medical attention as well.

Like the changes in eye wear prescription that many people undergo as they get older, our understanding of the ethics of human experimentation has gone through a number of evolutionary changes. This phenomenon is illustrated by the shifting ethical concerns regarding the selection of subjects for clinical research. Sequentially, each of these

views has held a given ethical problem to be central to the ethics of subject selection; in turn, each understanding of the primary ethical problem entailed an approach to the review of human subject research. I believe that the history of the ethics of subject selection in human experimentation can productively be divided into three periods.

It is not the case, though, that earlier concerns have fallen by the wayside. Each of the sequentially voiced ethical concerns continues to be relevant to the ethical analysis of research. Indeed, over the last decades the scope of ethical analysis regarding the selection of research subjects has broadened, each prescription change allowing us to see the world more clearly.

The 1970s: equitable distribution of burden

"Every family", observed the character Michael Corleone in the Godfather saga, "has skeletons in its closet". Were this proverb to be applied to human experimentation, it would have to be admitted that research abides in a house with many closets -- big, walk-in closets. The early history of research ethics was shaped by a number of prominent research scandals. Indeed, research ethics has been described -- and I think accurately -- as having been "born in scandal and reared in protectionism".[6] The Nuremberg

Code, often cited as the birth document of research ethics, was a response to the horrific experimentation undertaken by the Nazis on Jews, gypsies, Russians and other political prisoners during the Second World War.[7] Despite the existence of the Nuremberg Code and the later developed Declaration of Helsinki, the conduct of human experimentation in the United States remained relatively free of external regulation until the mid-1960s.

The perceived need for greater regulation of research in the United States was sparked by the revelation of an unsavoury experiment conducted in Brooklyn, New York. In 1963, three physicians at the Jewish Chronic Disease Hospital injected 22 chronically-ill patients with live cancer cells.[8] The purpose of the experiment was to determine if the lack of immune response against cancer cells observed in cancer patients was due to the fact that they had cancer or that they were chronically ill. Although the investigators had reason to believe that the patients in the study would "reject" the cancer cells, they could not be sure. Problematically, the research subjects were neither informed of the purpose of the experiment nor of the fact that the injections contained cancerous cells. In the wake of the public uproar resulting from the experiment, the Public Health Service asked the National Advisory Health

Council to "explore the advisability of establishing guidelines for the conduct of human experimentation in its extramural project-grants program".[9] Acting on the recommendations of the Council, Surgeon General Stewart issued a directive in 1966 indicating that the Public Health Service would not fund research unless it had been reviewed by an independent committee of peers.[9]

Formal regulations for the conduct of human experimentation in the United States were not instituted until the revelation of yet another scandal, the Tuskegee Syphilis Study. In 1972, it came to light that the Public Health Service had sponsored a study of 40 years duration in which 400 African-American men with syphilis were left untreated.[10] The study began in 1932, a time when the treatment for syphilis was both toxic and not very effective, and sought to determine whether the natural course of syphilis was, as was then thought, more benign in Afro-Americans. When penicillin -- a highly effective and non-toxic treatment for syphilis -- became available in the late 1940s, it was not offered to the study participants. Furthermore, study participants were actively deceived as to the nature of some study interventions. Spinal taps done purely for research purposes, for example, were described to participants as "treatments". As a result of the ensuing

scandal, Merlin Duval, Assistant Secretary of the Department of Health, Education and Welfare (DHEW) struck an Ad Hoc Advisory Panel to examine the Tuskegee Study and to examine the need for federal regulations to govern research.[9] As a result, in part, of the Panel's deliberations, the first Policy for the Protection of Research Subjects was issued by the DHEW in 1974.

Against this background of scandal and deceit, it is not surprising that the research ethics literature of the 1970s tended to characterize participation in clinical research as a risky endeavour. For example, Hans Jonas, in his classic essay, "Philosophical reflections on experimenting with human subjects" (1970), refers to participation in research as a "sacrifice" and to research participants as "martyrs".[11] Beginning with the premise that research participation is perilous, Jonas reasons that physician-researchers themselves ought to be the first subjects of human experimentation. If necessary, the scientists qua research subjects could be supplemented with other members of society beginning with the most educated and prosperous. Thus, Jonas' scheme for subject selection affords the most "captive" members of society -- the uneducated, the impoverished -- the greatest protection from the potential harms of research. For Jonas, the only way to

make the selection of subjects for research just is a schema that protects the disadvantaged from harm:

> [a]n inversion of the normal 'market' behaviour is demanded here -- namely, to accept the lowest quotation last (and excused only by the greatest pressure of need); to pay the highest price first.[11]

The National Commission shared Jonas' concern regarding the risks inherent in research participation. Consequently, they too were preoccupied with the issue as justice qua the equitable distribution of burdens in the selection of research subjects. For example, recommendation 4B in the National Commission's Report and Recommendations: Institutional Review Boards (1978) states that the IRB must ensure that the "selection of subjects is equitable".[12] Although this recommendation in itself could refer to the equitable distribution of either burden or benefit, the commentary that follows renders their fixation on burden perspicuous:

> [Comment on recommendation 4B] The proposed involvement of hospitalized patients, or other institutionalized persons, or disproportionate numbers of racial or ethnic minorities or persons of low socioeconomic status should be justified.[13]

The National Commission was motivated by past abuses to ensure that vulnerable subjects would be protected from human experimentation. The connection between the research scandals of the past and issue of justice in the selection

of subjects is explicitly drawn in the Belmont Report (1979). After a listing a bestiary of past abuses (including some of the cases discussed above), the members of the National Commission conclude that

> [a]gainst this historical background, it can be seen how conceptions of justice are relevant to research involving human subjects. For example, the selection of research subjects needs to be scrutinized in order to determine whether some classes (e.g., welfare patients, particularly racial and ethnic minorities, or persons confined to institutions) are 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.[3]

The notion that classes of subjects need to be protected from research is one that continues through to the most recent regulations for the Protection of Human Subjects from the Department of Health and Human Services (1991). Indeed, the sum total of the regulations' comments regarding the selection of subjects is as follows and seems to reflect this preoccupation with the equitable distribution of burden:

> [§46.111(a)] In order to approve research covered by this policy the IRB shall determine that all of the following requirements are satisfied:
> ...
> (3)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.[14]

It should be noted that category of so-called "vulnerable populations" is not only composed of those (e.g. pregnant women) who may be unduly susceptible to harm from research participation. Indeed, the majority of groups listed in this category are those who may be unable to give free and informed consent to research participation. Thus, the concept of equity in the distribution of the burden of research participation remained closely tied to concerned related to the principle of respect for persons. Levine in his book Ethics and Regulation of Clinical Research explains the connection as follows:

> [B]ecause we define as vulnerable those persons who are relatively or absolutely incapable of protecting their own interests through negotiations for informed consent, in the practical area of IRB review, there is often an interplay between considerations of informed consent and selection of subjects. To the extent that the subject population can be made less vulnerable in the sense of becoming more capable of protecting their own interests, fewer procedures are needed to assure the validity of their consent.[15]

Although little empirical information regarding the activities of Institutional Review Boards is available, it seems that their actions in the 1970s were in keeping with the protectionist tenor of the times. The most extensive survey of IRBs is contained within the National Commission's Appendix to Report and Recommendations: Institutional Review Boards.[16] Although IRBs tended to be much more preoccupied with issues of consent, approximately 4% of research protocols reviewed by IRBs had changes made to procedures for the selection of subjects. In fully 75% of these cases the changes required limited the scope of the study's sample; no case of an IRB requiring that subjects be added to a study is reported. The end result of the protectionist stance was that riskier research tended, by and large, to be done on middle aged males who reported high or middle annual incomes.[17]

The 1980s: equitable distribution of benefit

The turn of decade heralded a dramatic change in the public's perception of clinical research. In the span of a few short years, research, once perceived as risk laden, came to be seen as a source of potential benefit. Levine:

> 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.[18]

While the emergence of HIV/AIDS and AIDS activism certainly had much to do with the change in public attitude, a number of events prior to the advent of HIV/AIDS set the scene for change.

The thalidomide disaster in the early 1960s had a substantial impact both upon public perceptions of the potential hazards of pharmaceutical agents and on the regulation of new drug approval. In 1962 United States Congress passed the Kefauver-Harris amendments to the drug approval laws which required for the first time that the Food and Drug Administration (FDA) ensure that drugs were not only safe but also effective before they could be licensed. By the late 1970s, though, objections were being voiced by industry to the inordinate time delays involved in the drug licensure process.[19] Eventually, in 1983 FDA modified some of its procedures to attempt to speed the drug approval process.[19]

Coincident with these events, a series of studies was published (1976 to 1982) which attempted to quantify the risk to subjects conferred by research participation.[20,21,22,23] Overall, the studies indicated that participation in research was a relatively safe activity. Non-therapeutic

trials, including phase I studies to test the safety of new drugs in humans, seem to pose the least risk to participants.[21] Therapeutic studies, including phase II and phase III clinical trials, tended to pose more risk to subjects than non-therapeutic studies.[23] Even so, the majority of the serious adverse events seemed to be related to toxic therapy, such as chemotherapy for persons with cancer, which posed risks to subjects that were similar (at least in kind) to those present in clinical practice.[20]

Clearly, though, HIV/AIDS had the most substantial impact on the public's perception of clinical research. Although the disease was first recognized in 1981, no antiviral therapy was tested for efficacy against the disease until 1986. Thus, for five years, no proven treatment existed for HIV/AIDS. The first randomized controlled trial of zidovudine (AZT) was a phase II, placebo controlled trial: 145 subjects received AZT, while 137 received placebo.[24] Most of the subjects were white, homosexual men; all had either AIDS or ARC (AIDS-related complex; a pre-AIDS syndrome); and, intravenous drug abusers were explicitly excluded from study participation. Although accrual to the study was completed on June 30, 1986, the study was stopped only three months later. By September, the advantage conferred to subjects treated with AZT was

obvious: 16 subjects had died on the placebo arm compared to only one treated with AZT.

As the results of the AZT trial became widely known, persons with HIV called for access to therapy with AZT (or other promising treatments) but, were unable, in many cases, to obtain it outside of clinical trials. Persons with HIV/AIDS responded vocally and called for access to experimental treatments. Groups such as ACT-UP ("AIDS Coalition to Unleash Power") staged protests that were effective in raising the public's awareness of the issue. The desire of patients to gain access to the potentially life-saving treatments in trials was great: in some cases, patients (occasionally with the aid of their physicians) falsified medical data to satisfy the criteria for trial eligibility.[19]

Participation in research had, at times in the past, been seen as a benefit. In the early debate on the permissibility of including prisoners in clinical research, some prisoners voiced the opinion that research participation was a benefit -- monetary rewards for participation exceeded other prison jobs and, perhaps more importantly, it represented an opportunity to be altruistic. The shift that occurred with HIV/AIDS, though, was that some groups in society began to see trial participation as

essential to their own medical care and, ultimately, to their own chances for survival.

The events of the 1980s shaped the nature of the discourse of research ethics at the time. Carried along by the events of the day, ethicists too came to see access to trials as the prime issue in the selection of subjects. The issue remained one of justice but the emphasis had shifted from equitable distribution of burden to the equitable distribution of benefit. As a result, ethicists began to question the exclusion of groups of patients from clinical trials. For example, writing in 1986, Macklin and Friedland write:

> As well as the problem of fair distribution of the burdens of research, the issue of justice in AIDS research also seems to include the opposite problem: Who will receive the benefits of early testing of promising new drugs? The problem is already apparent in the phase II study of AZT, which was performed almost entirely on homosexual male patients, excluding intravenous drug abusers. Although drug abusers were excluded according to a 'medical' rationale -- this group tends to be 'unreliable' and 'noncompliant', and hence is not a good study population -- the resulting distribution of benefits was nonetheless unjust.[24]

When participation in research came to be seen as a benefit, the exclusion of potentially vulnerable groups from research fostered by the protectionism of the 1970s came to be seen itself as an injustice. For example, Carol Levine, writing in 1988, asks:

> How can groups of prospective subjects traditionally excluded from clinical trials because of their physical or social vulnerability (women of childbearing age, infants prisoners, intravenous drug users, prostitutes) be given access to clinical trials that may, perhaps, prove of benefit to them?[24]

Thus, towards the latter part of the decade and into the early 1990s, researchers and IRBs alike were advised that the exclusion of groups of affected individuals would have to be carefully justified. For example, consensus statement 7 from Carol Levine et al.'s "Building a new consensus: ethical principles and policies for clinical research on HIV/AIDS" stated 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.[19]

Although such requirements have yet to make their way into the Department of Health and Human Services Regulations (cited supra), the Institutional Review Board Guidebook published by the Office for Protection from Research Risk does require IRBs to consider the following points when reviewing research:

- "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 disadvantage persons, patients of researchers, seriously ill persons) so that they are denied opportunities to participate in research?".[25]

But, do IRBs change or eliminate eligibility criteria on the basis of equitable distribution of benefits? To the best of my knowledge, no recent, comprehensive empirical study of the activity of IRBs has addressed this issue. Freedman recently reviewed the actions of a single committee over a two year period (February 1990 to January 1992).[26] Of 191 protocols approved by the committee, twenty-five protocols had at least one eligibility criterion changed by the committee. According to Freedman, in five cases (20%) an eligibility criterion requiring an HIV test was dropped. My own experience on that same committee (1994) and another IRB (1993 to present) -- encompassing the review of perhaps 300 protocols -- leads me to believe that it is not rare for committees to question eligibility criteria that exclude certain groups -- particularly women of reproductive age, the elderly, persons with a history of drug or alcohol

abuse, and otherwise healthy persons with HIV -- who might benefit from research participation. A systematic study of the actions of IRBs in this regard is required to answer this question definitively.

The 1990s: applicability of the results of research

In the 1990s, a new issue has been added to the ethical debate regarding the inclusion of groups hitherto excluded from research, namely, that if the results of research are to be widely applicable, i.e., maximally beneficial to society, then subjects included in research studies must be representative of the population of persons affected. Constructed as such, the issue is beneficence -- social benefit -- rather than justice.

Concern in the ethics literature regarding the applicability of narrowly focussed studies to the population at large can be traced to the early HIV/AIDS literature. For example, Macklin et al (1986) points 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. For example, few women and no children or intravenous drug abusers were included in the study.[24]

A relatively muted voice in the early ethical debates regarding subject selection, a call for broader inclusion of

subjects based on concerns regarding the applicability of study results receive equal attention with justice-based arguments in Carol Levine et al.'s "Building a new consensus: ethical principles and policies for clinical research on HIV/AIDS" (1991). For example, in arguing for a presumption of inclusiveness in the selection of subjects for trials the authors argue that

> [i]t is essential that data be developed to serve the well-being of all groups affected by the research [beneficence]...In addition, it is necessary to assure equitable access to clinical trials for all affected groups within communities [justice].[19]

In the same year (indeed in the same issue of IRB), Carol Levine drew the connection between the existence of possible biological differences in the way HIV/AIDS affects women and inability to extrapolate treatment data derived from studies of men to the treatment in clinical practice of women with HIV/AIDS.[27] Given these biological differences, she argues, studies would need to enroll sufficient numbers of women to address separately questions such as the efficacy of a drug and the drug's optimal dose (for women). When these issues are not adequately addressed in clinical studies Levine claims that women suffer because practitioners have insufficient information to treat them properly. Not only must barriers to enrollment of women in

studies be removed, but women must be adequately represented in studies. This, she observes, will require

> [r]ecruitment efforts...[that] take account of the multiple roles HIV-infected women play as family care givers and employees (often in marginal jobs with few opportunities for flexibility). Meeting their own health care needs may not be their highest priority; enrolling in research, an alien concept to many, may seem less important.[28]

Starting in 1992 and continuing to the present, the issue of the applicability of research results dramatically increased in scope. In an influential article in the Hastings Center Report in 1992, Rebecca Dresser argued that the exclusion of women from trials pertained to more than HIV/AIDS research: "the failure to include women in research populations", she observed, "is ubiquitous".[28] Indeed, according to Dresser, not only women, but also racial and ethnic minorities are largely underrepresented in research designed to establish the efficacy of new medical treatments. Pointing to the biological differences that may exist between genders and amongst racial and ethnic groups, she concludes that "[s]uch differences make it inappropriate to generalize findings based on one gender or racial group to all human beings".[29] Thus, optimal clinical trials will include representative numbers of both women and racial/ethnic minorities. The main ethical issue at stake is, for Dresser, beneficence:

> The current disparity between the health information we have about white males and the information we have about women and people of color contravenes basic ethical principles governing human experimentation. Most clearly violated is the principle of beneficence, which holds that biomedical research should be designed to maximize benefit and minimize harm.[29]

Following Carol Levine, Dresser notes that special recruitment efforts may be needed to redress the underrepresentation of these groups in clinical research.

> If scientists conducting [for example] heart disease research are afraid that physicians will not refer enough female patients, then extra recruiting measures should be taken to ensure that they do. Research costs may go up, but the benefits of including women subjects [and minorities] are worth the expense.[29]

In 1993, politicians entered the fray. In that year, the U.S. Food and Drug Administration released its "Guidelines for the study and evaluation of gender differences in the clinical evaluation of drugs". The new guidelines withdrew the Agency's long-standing prohibition on the participation of women of childbearing potential in early clinical trials. Ethical arguments regarding the applicability of research results clearly influenced the new FDA policy. Although the new Guidelines did not require the inclusion of women in studies in all cases, the Guidelines state that

> The patients included in clinical studies should, in general, reflect the population that will receive the drug when it is marketed. For most

> drugs, therefore, representatives of both genders should be included in clinical trials in numbers adequate to allow detection of clinically significant gender-related differences in drug response...Such analyses of subsets with particular characteristics can be expected to detect only relatively large gender-related differences, but in general, small differences are not likely to be clinically important.[29]

The change in FDA policy reflects a shift from policy grounded in non-maleficence -- avoiding harm -- to a policy concerned with beneficence, that is, ensuring the wide-spread applicability of research results. The absence of reference in the Guidelines to arguments based on the equitable distribution of benefit is noteworthy.

Perhaps even more significant than the changes to the FDA regulatory policy was the signing into law in June 1993 of the "Clinical Research Equity Regarding Women and Minorities" provision of the NIH Revitalization Act. The law required the Director of NIH to construct guidelines to ensure that both women and minorities are included as subjects of NIH funded research. The new NIH Guidelines, published in 1994, require that

> [T]he NIH must:
> - Ensure that women and members of minorities and their subpopulations are included in all human subject research.
> - For phase III clinical trials, ensure that women and minorities and their subpopulations must be included such that valid analyses of differences in intervention effect can be accomplished;

> - Not allow cost as an acceptable reason for excluding these groups; and,
> - Initiate programs and support for outreach efforts to recruit these groups into clinical studies.[30]

As with the FDA Guidelines, the new NIH Guidelines were justified with arguments based on the applicability of research results:

> 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 behavioural 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.[31]

Clearly, these changes to the NIH and FDA regulations will have implications for the design of trials and their review by Institutional Review Boards. As required by justice-based concerns, an IRB must carefully scrutinize the justifications for eligibility criteria that deny access to population groups to clinical trials. Beneficence, and in particular the issue of the applicability of the results of trials, implies additional duties. (These are outlined in a letter to IRBs by Gary Ellis, Director of the Office for Protection from Research Risk, April 25, 1994). For NIH

funded research, IRBs have to ensure that adequate representation of certain groups within the trial is likely. This certainly entails scrutinizing proposed eligibility criteria for studies. This also entails that the IRB: (a)review procedures for subject recruitment, particularly the recruitment of subjects from hard-to-access groups including intravenous drug users and African-Americans; (b)require the development and review of programs to attempt to ensure the continued participation of subjects in the study; (c)finally, IRBs may require, as a part of their annual review process, that investigators provide them with demographic information regarding accrual to ensure that these procedures are effective. As these changes have just taken effect in the 1995 fiscal year, no empirical data are available to document the actual role that IRBs are playing in this regard.

We have seen that a succession of issues have been viewed by ethicists to be the prime issue in the selection of subjects for clinical research. Each of these changes in focus has been the result of a complex interaction among current events (e.g. the revelation of scandal), political factors (e.g. advisory panels, new regulations) and ethicists themselves (e.g. Dresser's article undoubtedly influenced the political debate). Overall, the most dramatic

shift surrounding the selection of subjects for trials has been the shift from "excluding" the vulnerable to a presumption to include persons from all groups. This change has paralleled a shift in the public's perception of participation in trials from a risky activity to one that possibly offers benefit.

In dividing the ethical concerns into three issues and in assigning these issues to decades, we have, no doubt, done some violence to the true complex interplay of these issues. Indeed, all three of the issues -- equitable distribution of burden, equitable distribution of benefit, and the applicability of research results -- currently occupy the thoughts and activities of ethicists and IRB members. Each new area of ethical concern in the selection of subjects added to the range of ethical inquiry; in turn, the scope of the review of protocols by IRBs progressively enlarged.

This narrative has described two general trends. First, that concerns of research ethicists (and IRBs) regarding the selection of subjects for research has shifted in emphasis over the last three decades. Second, that the scope of ethical inquiry regarding selection of subjects has broadened over the same time; first including the protection of subjects, then expanding access to clinical trials, and

finally, encompassing the issue of the applicability of the results of trials to clinical practice.

Study questions

Given these recent developments in the regulation and ethics of clinical research, the impact of eligibility criteria on the generalizability of randomized controlled trials represents an important area of inquiry. Rather than a strictly philosophical or historical approach to the issue, I propose to undertake an approach that is partly empirical. Thus, in the following two chapters I will address the following questions:

- Is a large proportion of patients with a given disease treated in clinical trials?
- Are subjects randomly selected from the patient population for participation in clinical trials?
- Are clinical trial participants comparable to non-participants?

Each of these questions bears on the generalizability of the results of clinical trials. For each of the study questions an affirmative answer is consistent with the conclusion that the results of clinical trials are widely generalizable.

The final chapter of the thesis will review the findings presented in chapters two and three and will place those findings within the context that we have laid out here. A final section will address the implications of this research for the review of clinical research by ethics boards -- REBs or IRBs.

References

1.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.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.

3.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.

4.Beauchamp T. Distributive justice and morally relevant differences. In: The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Appendix Volume I The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Washington, D.C.: Department of Health, Education and Welfare, 1978 (DHEW Publication No. (OS) 78-0013). Pp. 6.1-6.20.

5.Wittgenstein L. Philosophical Investigations. Trans. G.E.M. Anscombe. Oxford: Basil Blackwell Ltd., 1991: §103.

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

7.Annas GJ, Grodin MA (eds.). The Nazi Doctors and the Nuremberg Code: Human Rights in Human Experimentation. New York: Oxford University Press, 1992.

8.Katz J. Experimentation With Human Beings. New York: Russell Sage Foundation, 1972: pp. 9-65.

9.Katz J. The regulation of human experimentation in the United States -- a personal odyssey. IRB: A Review of Human Subjects Experimentation 1987; 9(1): 1-6.

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

11.Jonas H. Philosophical reflections on experimenting with human beings. In: Philosophical Essays: From Ancient Creed to Technological Man. Chicago: University of Chicago, 1974:

pp. 105-131.

12.The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Report and Recommendations: Institutional Review Boards. Washington, D.C.: Department of Health, Education and Welfare, 1978 (DHEW Publication No. (OS) 78-0008). Page 19.

13.See *supra* note 12, p. 22.

14.Department of Health and Human Services. Protection of Human Subjects. Tile 45, Code of Federal Regulation, Part 46: Revised June 18, 1991.

15.Levine RJ. Ethics and Regulation of Clinical Research (2nd edition). Baltimore: Urban & Schwarzenberg, 1986: 72.

16.The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Appendix to Report and Recommendations: Institutional Review Boards. Washington, D.C.: Department of Health, Education and Welfare, 1978 (DHEW Publication No. (OS) 78-0009): page 1.120.

17.See *supra* note 12, pp. 64-65.

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

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

20.Cardon PV, Dommel FW, Truble RR. Injuries to research subjects: a survey of investigators. New England Journal of Medicine 1976; 295: 650-654.

21.Zarafonetis CJD, Riley PA, Willis PW, et al.. Clinically significant adverse effects in a phase one testing program. Clinical Pharmacology and Therapeutics 1978; 24: 127-132.

22.McCann DJ, Pettit JR. 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.

23.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 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.

24.Macklin R, Friedland G. AIDS research: the ethics of clinical trials. Law, Medicine and Health Care 1986; 14: 273-280.

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

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

27.Levine C. Women and HIV/AIDS research: the barriers to equity. IRB: A Review of Human Subjects Research 1991; 13(1-2): 18-22.

28.Dresser R. Wanted single, white male for medical research. Hastings Center Report 1992; January/February: 24-29.

29.Food and Drug Administration. Guidelines for the Study and evaluation of gender differences in the clinical evaluation of drugs. Federal Register 1993; 58: 39406-39416.

30.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.

Chapter 2:

Eligibility criteria and other barriers to enrollment in randomized controlled trials

Abstract

Introduction. Despite the central role of the randomized controlled trial in the genesis of medical knowledge, only a small proportion of patients are treated in RCTs. This fact may have implications for the time needed to complete trials and for the generalizability of the results of trials to clinical practice.

Methods. A review of empirical studies was undertaken in order to quantify the relative contributions of eligibility criteria, physician refusals and patient refusals to the failure to enroll subjects in RCTs. Empirical studies reporting on the enrollment of subjects to RCTs were retrieved from the Medline (1966-1994) and Cancerlit (1983-1994) databases. The number of subjects excluded due to eligibility criteria, physician refusal, patient refusal or other factors was recorded.

Results. Criteria for trial eligibility proved to be the largest barrier to trial enrollment. Overall, 53% of subjects for whom a trial was available for their type and stage of disease were ineligible for trial participation. Physician refusal and patient refusal each accounted for exclusion of 7% of the potential subjects (1% were excluded for other reasons).

Discussion and Conclusion. Since many eligibility criteria, particularly in North American RCTs, may be unnecessary, trial designers ought to hold broad-based, pragmatic trials as an ideal. Furthermore, individual eligibility criteria, when included in a clinical trial protocol, ought to be justified to highlight their necessity. The relaxation of criteria for clinical trial enrollment, when combined with other approaches, ought to increase the proportion of patients treated in RCTs.

Introduction

The randomized controlled trial (RCT) is the gold standard in the evaluation of the safety and efficacy of novel medical interventions. Only a small proportion of cancer patients, however, are enrolled in RCTs. Tate et al. reviewed enrollment of cancer patients in RCTs in the United Kingdom and found that only 3.7% of patients were treated in trials.[1] Friedman et al. reviewed enrollment in National Cancer Institute (NCI) funded trials in the United States and found that only 1.6% of cancer patients were treated in phase II and III clinical trials.[2] Since cancer is one of the diseases most actively studied by clinical research, it is likely that only a small proportion of other patient populations are enrolled in randomized controlled trials.

The enrollment of a small proportion of patients in RCTs is not, in itself, a problem. So long as RCTs can enroll subjects at a sufficient rate to answer efficiently clinical questions and so long as subjects enrolled in RCTs are reasonably representative of patients in clinical practice, it matters little what proportion of patients is treated in trials. Unfortunately, both of these conditions may be lacking. The NCI has reported that many of its RCTs have rates of enrollment that are considerably slower than the planned rates.[3] This means that periods for subject

enrollment for RCTs may last from 3 to 5 years or even longer.[3] Enrolling a larger proportion of patients in RCTs could solve this problem.

Regarding the second condition, Begg et al. have pointed out that criteria for clinical trial eligibility are frequently so restrictive as to call into question the widespread applicability, or generalizability, of the results of RCTs.[4] The enrollment of a larger proportion, or at least a more representative proportion, of patients to RCTs is required.

Why are so few subjects enrolled in randomized trials? Protocol factors, physician factors and patient factors have all been cited as contributing to the low proportion enrolled.[4,5,6] In 1991, Gotay reviewed empirical studies that examined the relative contribution of each of these factors to exclusions from RCTs.[7] Gotay found that "fewer than half of the available patients were eligible for a particular clinical trial".[7] She also concluded that physician and patient factors added substantial losses to trial enrollment.[7]

Gotay's review, however, has serious flaws. We are not told how the studies were retrieved for the review, nor are all of the relevant studies incorporated therein. More fundamentally, when labelling subjects as "ineligible",

Gotay fails to separate two groups of subjects. The first group is comprised of subjects for whom no RCT is available for their type and stage of disease. The second, and more important group, is comprised of subjects for whom a trial is available but who fail to fulfil the criteria for clinical trial eligibility. Since the results of a RCT potentially apply to subjects in the second group but not to those in the first, it is only the exclusion of subjects in the second group that brings the generalizability of the trial's results into question. A systematic review of the empirical literature is thus required to assess and quantify the relative contributions of protocol, physician and patient factors to enrollment losses in RCTs.

Methods

Published reports of empirical studies were retrieved through searches of the Medline and Cancerlit databases for all available publication years (i.e. Medline: 1966 through October, 1994; Cancerlit: 1983 through September, 1994). The following keywords were used to define the searches: "clinical trials" and "eligibility determination", "patient participation" or "registries". The searches were supplemented with articles cited in the bibliographies of relevant papers. The names of the authors, publication year,

study population and sample size were recorded from each study. Subjects of trials noted in these studies were then divided into one of the following mutually exclusive and exhaustive categories:

(1) Subject enrolled in a RCT or,

Subject not enrolled in a RCT due to:

(2) the absence of a protocol for his or her type and stage of disease;

(3) failure to meet the criteria for trial eligibility (but not due to 2);

(4) refusal of his or her physician to enroll them (but not due to 2 or 3);

(5) subject refusal to consent to study participation (but not due to 2, 3 or 4, unless ineligibility is solely due to refusal to consent); or,

(6) other (or unknown) reasons (i.e. not due to 2, 3, 4 or 5).

In most cases, the actual number of subjects in each category was reported in the study. In some cases, numbers of subjects had to be reconstructed from reported percentages. In a few cases, information on one or more categories was not available. Studies were included in this review if they reported at least the number of subjects

excluded due to criteria for clinical trial eligibility.

Studies were divided into three groups: (1) studies that assessed a population cohort for eligibility for two or more randomized controlled trials (hereafter, group 1); (2) studies that assessed a population cohort for eligibility for a single RCT (group 2); and, (3) studies reporting patient logs from individual RCTs in which subjects with a given type and stage of disease were assessed for eligibility (group 3). To ensure that the mean proportions summed to unity, only studies that reported information on all categories were included in the calculation of descriptive statistics.

Proportions were calculated using the number of subjects for whom a trial was available as the denominator. This approach allowed for the possibility of combining results across groups since group 3 studies only assessed subjects for whom a trial was available. Furthermore, for the reason outlined above, this denominator is the most relevant to the assessment of the generalizability of trial results. Both unweighted and weighted (for sample size) mean proportions were calculated for each study group. It was decided <u>a priori</u> that overall (across groups) estimates would be calculated only if the results from each of the groups were comparable.

Results

The review of the literature retrieved twenty-two studies from twenty articles that document the eligibility of 148,561 subjects for randomized controlled trials. The RCTs in question addressed the treatment of patients with cancer, heart disease, affective disorders and stroke. Nine studies (from eight articles) fell into group 1 (N=69,323) and five studies fell into group 2 (N=26,070). It is likely that the search strategy retrieved from the literature all or most of the existing studies in these first two groups. Eight studies (from seven articles) documenting the eligibility of 53,168 subjects fell into group 3. As RCTs are reporting results from patient logs with increasing frequency, these eight studies very likely represent only a small proportion of such studies in the literature. Seventeen of the twenty-two studies reported information on all of the categories and were thus included in the calculation of descriptive statistics. In all, these seventeen studies document the enrollment experience of 51,736 subjects for whom a clinical trial was available.

The data from the studies in group 1 are reported in table 1.[8,9,10,11,12,13,14,15] A large percentage (37%) of the subjects assessed for RCT enrollment had no trial available to them. Of the subjects for whom a trial was available,

only 36% were finally enrolled in a RCT. Failure to fulfil criteria for clinical trial eligibility excluded the largest proportion (45%) of subjects for whom a trial was available, followed by physician refusal (9%), patient refusal (9%) and other reasons (1%). Weighted means are reported and provide similar results.

The data from the studies in group 2 are reported in table 2.[16,17,18,19,20] Of the subjects assessed in these studies, 55% had the "wrong" stage of disease and thus had no trial available to them. Of those for whom a RCT was available, 29% were enrolled. 58% of subjects for whom a trial was available were not enrolled due to ineligibility, 7% due to physician refusal, 4% due to patient refusal and 2% for other reasons. The weighted mean proportions revealed similar results.

The data from the studies in group 3 are reported in table 3.[21,22,23,24,25,26,27] In total, 30% of the subjects were enrolled in a RCT. 56% of the subjects for whom a trial was available were excluded due to ineligibility, 4% were excluded due to physician refusal, 9% were excluded due to patient refusal and 1% were excluded for other reasons. Weighted means were comparable to the unweighted means reported.

Since the descriptive statistics were similar in all

three groups, they were combined into summary descriptive statistics. These summary statistics are presented in figure 1. The unweighted means are as follows: of subjects for whom a trial was available, 32% were enrolled in a randomized controlled trial, 53% were ineligible for trial participation, 7% were excluded due to physician refusal, 7% refused to consent to participate, and 1% were excluded for other reasons. Weighted means were similar to these reported values (35%, 47%, 9%, 6%, and 3%, respectively).

Finally, substantial differences were noted between studies based in North America and those based elsewhere. (Of the seventeen studies with complete results, all were conducted either within North America or elsewhere; that is, none were conducted both within and outside of North America: so the above categories do not overlap). In the nine North American studies, only 19% (weighted mean) of subjects for whom a trial was available were enrolled in a RCT. 55% of those for whom a trial was available were excluded due to ineligibility, 15% were excluded due to physician refusal, 9% were excluded due to patients refusal, and 2% were excluded for other reasons. In the eight studies not based in North America, 58% (weighted mean) of subjects for whom a trial was available were enrolled in a trial, nearly three times the proportion enrolled in North American

studies. 36% of subjects for whom a trial was available were excluded due to ineligibility, 0% were excluded due to physician refusal, 3% due to patient refusal, and 3% for other reasons.

Discussion

The observant reader will have noted a discrepancy in our figures. According to our literature review, 32% of patients for whom a trial is available are enrolled in clinical trials. But the actual proportion of cancer patients enrolled in RCTs, as noted at the start of the paper, is one-tenth of this figure. Why is this so? First, many disease types and stages have no RCT available for them. In our first group of studies, 37% of incident cases had no trial available. Second, in other cases, classes of patients may not themselves be "available" to be considered for enrollment in a RCT. One class of patients who are potentially "unavailable" are those treated outside of tertiary-care centres. Efforts to extend RCTs to include community hospitals have proven an effective way of increasing trial enrollment.[28,29,30] Another class of "unavailable" patients are those whose physician is unaware of the existence of individual RCTs or may be reluctant to enter patients (in general) in trials for a variety of

reasons.[5] In response to this, the NCI's Office for Cancer Communications has initiated a program to heighten the awareness of both physicians and patients of individual clinical trials.[3] If enrollment in RCTs is to be maximized, though, barriers to the enrollment of subjects who have a trial available for them will need to be addressed.

This review demonstrates that criteria for clinical trial eligibility are the largest barrier to RCT enrollment for subjects who potentially have a trial available to them (and who themselves are available for trial participation). Physician refusals and patient refusals each account for less than one-seventh of the proportion of subjects excluded by eligibility criteria. But are any of the enrollment losses due to eligibility criteria avoidable? Put another way, are all of the eligibility criteria in clinical trials necessary?

There are several indirect lines of evidence which suggest that some eligibility criteria may be unnecessary. Begg et al. reviewed eligibility criteria in nine concurrent RCTs for the treatment of breast cancer.[4] They found that the trials studied contained many eligibility criteria, some of which were seemingly arbitrary. Substantial variation among trials was noted in criteria that defined the maximum time allowable since surgery, acceptable values for tests of

organ function and concomitant disease exclusions. Begg et al. concluded that "the variation [in criteria] observed represents a lack of consensus on the need for specific restrictions. Moreover, the greater the variation across studies, the more we must be skeptical about the value of a particular restriction".[4]

Our own comparison of a twenty-year sample of breast cancer trials from the National Surgical Adjuvant Breast Project (NSABP) with a set of acute lymphocytic leukaemia (ALL) trials from the Pediatric Oncology Group (POG) provides further information.[31] Although the number of criteria doubled in both trial groups over the time period, the number of criteria in the POG trials (6 to 12) remained substantially fewer than the number of criteria in the NSABP trials (22 to 44). Perhaps as a result of differences in defining eligible subject populations in the two cancer types, the proportion of patients with ALL enrolled in RCTs (79%)[32] is much greater than the proportion of breast cancer patients enrolled in trials (3.3-8%).[1,2]

Finally, the discrepancy between North American studies and non-North American studies observed in this review may indicate that North American RCTs have eligibility criteria that are unnecessary. North American studies excluded 55% of subjects due to ineligibility while studies conducted

elsewhere excluded only 36%. Some of the non-North American studies appear to have used very few eligibility criteria indeed. Anderson reports that three RCTs of the Danish Breast Cancer Cooperative Group studying the treatment of breast cancer with adjuvant therapy contained a minimal set of criteria ("operable breast cancer with no metastases", "no medical contraindication to the study treatments" and "age less than 70").[13]

The inclusion of "no medical contraindication" as an eligibility criterion may explain why no subjects are excluded due to 'physician refusal' in the non-North American studies. But isn't this single criterion really a proxy for a whole set of criteria made explicit in North American protocols? Yes and no. Insofar as North American protocols make explicit with eligibility criteria subsets of patients for whom the study treatments are contraindicated (i.e. those patients who would not be treated in clinical practice), these criteria are identical to the "no medical contraindication" criterion. These criteria are necessary (either in an abbreviated or an explicit form) and do not diminish the applicability of a RCT's results to clinical practice. North American trials, though, typically exclude further categories of patients -- patients with, for example, a history of cancer, or comorbid conditions -- from

trials: patients who would be treated in clinical practice. These criteria are not essential -- either to protect patients from harm or to make the population of patients more homogeneous -- and have the disadvantage of reducing the generalizability of a trial's results. Indeed, our above-mentioned comparison of NSABP and POG clinical trials revealed that the bulk of the increase in eligibility criteria over the twenty year period was accounted for by criteria designed to make the study population more homogeneous (criteria directed at patient safety remained constant).[31]

Why do some RCTs have unnecessary eligibility criteria? Begg et al. offer three answers to this question: (1) fear of excessive toxicity, (2) attempting to attain homogeneity and, (3) concern over qualitative interaction.[4] Let us discuss each of these in turn.

Fear of excessive toxicity. Particularly in trials with toxic treatments, such as cancer chemotherapy trials, concern over toxicity may motivate trial designers to exclude subjects thought to be unduly or unusually susceptible to harm, including those who are old (e.g. excluding persons >70 years of age), those with abnormal organ function (e.g. excluding persons with liver enzymes greater than 1.5 times the upper limit of normal) and those

with comorbid conditions (e.g. excluding patients with cardiac disease).[4] Clearly, subjects for whom one of the trial treatments is contraindicated must be excluded. But, as noted above, the necessity of excluding persons thought to be susceptible to harm from trials is uncertain: the risk associated with many "toxicity" exclusion criteria is ill-defined. The lack of information on risk is, indeed, largely due to the routine exclusion of these groups from RCTs.

The routine exclusion of patient subgroups by "toxicity" eligibility criteria has a number of detrimental effects. First, little information exists to guide practitioners in the treatment of persons in these subgroups. For example, as a result of the routine exclusion of persons over the age of 70 from cancer treatment trials, little information exists on the treatment of cancer in the elderly.[33] Second, unsuspected treatment benefits may remain undiscovered. Concerns over the risk of haemorrhagic complications from intravenous thrombolysis motivated most trials testing thrombolytics in the treatment of acute MI to exclude persons over the age of 75.[34] Results from the ISIS-2 trial indicate, however, that the benefit from thrombolysis may be greatest over the age of 75.[34] Finally, "toxicity" exclusions may have unintended effects in RCTs. In a review of 214 clinical trials of various therapies in

the treatment of acute MI, Gurwitz et al. found that 61% of the trials excluded persons over the age of 75.[35] Since more women than men survive past the age of 75, the trials with age-based exclusions had a lower proportion of women participants.[35]

Attempting to attain homogeneity. Begg et al. speculate that many trial designers have been motivated by the ideal of the "wet-bench experiment" which utilizes homogeneous experimental units.[4] Trial designers commonly attempt to increase the homogeneity of study population by excluding subjects thought to have differing prognoses. The laboratory ideal of homogeneous experimental units cannot, however, be achieved in the clinical setting where "between-patient variation is always large relative to the anticipated treatment effect".[4] Yusuf et al. point out that this unavoidable heterogeneity is due to the fact that "all variables that influence an outcome are not known and these unknown variables can have a substantial impact upon prognosis".[36]

The solution is not found in defining the eligible subject population narrowly. Restricting the eligible population will both diminish the ability of an RCT to enroll sufficient patients and will reduce the relevance of the RCT's results to clinical practice.[35] Rather, the answer

is to randomize large numbers of subjects so that those with differing prognoses will tend to be evenly distributed among the treatment arms in a RCT.[35,36]

What if circumstances seem not to allow for the randomization of larger numbers of subjects? What if funding or human resources or both are restricted? The combined approach of minimizing criteria for trial eligibility and simplifying other aspects of a RCT offers a relatively cost-free approach to increase trial enrollment. If eligibility criteria are minimized, the financial and human resource "cost" per patient is reduced: fewer investigations are required, less paper work needs to be filled out and the amount of physician time required is diminished.[37]

Qualitative interaction. As stated above, subgroups of patients in whom a therapy is known to be harmful must be excluded from a trial. However, additional groups of patients are frequently excluded from trials on the basis that the treatment effect may be in a different direction in that group (qualitative interaction) or in the same direction but of diminished magnitude (quantitative interaction).[4] It seems though that unexpected qualitative interactions are rare. Yusuf et al. examined widely-differing groups of patients treated with antiplatelet drugs, beta-blockers and calcium channel blockers and found

that the direction of effect was the same in all groups.[36] Furthermore, groups excluded because of concern over qualitative effects have, in some cases, subsequently been proven to have benefits similar in magnitude to those of other trial participants. For example, the concern that late reperfusion therapy might induce early mortality after acute myocardial infarction motivated a number of trials of thrombolytic agents to restrict trial eligibility to those who presented within 6 hours of the onset of symptoms. Yet, Yusuf's meta-analysis demonstrated that the treatment benefit in the 7-24h post-MI group was in fact comparable to that in the early presentation (< 6 h) group.[38] Similar concerns motivated the routine exclusion of patients with congestive heart failure from RCTs testing the effect of ß-blockers post-MI. It seems though that the reduction in mortality offered by beta-blocker therapy is similar in patients with and without heart failure.[36]

Even if quantitative interactions are known to exist for certain subgroups of patients, it may be more efficient to include them in a trial than to exclude them. When efficiency is measured in terms of trial duration, this is not difficult to understand. By excluding these subgroups, the pool of subjects eligible for the trial is reduced and, hence, the rate of enrollment in the trial is slowed. Thus,

gains in terms of enrolling only subjects likely to have a large positive outcome can be quickly lost to slow rates of enrollment in the RCT. Buyse reports the results of an armchair experiment in which the efficiency of including a poorer prognosis group of patients in a trial is calculated with varying degrees of qualitative interaction in the poor prognosis group and varying proportions of poor prognosis patients in the patient population as a whole.[39] Buyse found that including the poor prognosis group increases efficiency across a broad range of assumptions.[39]

It is clear from the results of this review that any attempt to improve the proportion of patients enrolled in randomized controlled trials ought to address the issue of subject eligibility. The most obvious and efficient solution would be for RCT designers to embrace the ideal of the broad-based, pragmatic randomized controlled trial. By mirroring clinical reality as closely as possible, a RCT will maximize the usefulness of its results to clinicians and thus, will optimize the impact of the trial on clinical practice. Furthermore, by utilizing a minimal set of eligibility criteria, the RCT will maximize the pool of eligible subjects available for trial enrollment and thus, decrease the duration of the enrollment phase of the trial. In keeping with the ideal of a minimal set of eligibility

criteria, when individual eligibility criteria are required, they ought to be individually justified to highlight their necessity. Critically addressing the issue of subject eligibility will, when used in concert with the approaches to the enhancement of enrollment outlined above, maximize patient enrollment in RCTs.

Table 1.

Study	Study population	Sample size	Reason for non-participation					Enrolled in trial
			No protocol available	Ineligible	Physician refused	Subject refused	Other exclusion or unknown	
			Number (percentage)					N (%)
Greco (1980)†	cancer patients	202	89 (44)	32 (16)	14 (7)	33 (16)	0 (0)	34 (17)
McCusker (1982)†	cancer patients	454	112 (25)	196 (43)	19 (4)	8 (2)	6 (1)	113 (25)
	lung cancer	347	81 (23)	108 (31)	?	?	49 (14)	31 (9)
Lee (1983)†	radiotherapy patients	1103	703 (64)	137 (12)	118 (11)	21 (2)	0 (0)	124 (11)
Begg (1983)†	cancer patients	3534	774 (22)	1665 (47)	287 (8)	105 (3)	66 (2)	637 (18)
Hunter (1987)†	cancer patients	44156	26383 (60)	9486 (22)	2859 (6)	1794 (4)	392 (1)	3242 (7)
Anderson (1988)	breast cancer	18487	1565 (9)	4827 (26)	0 (0)	234 (1)	504 (3)	11357 (61)
Jack (1990)	breast cancer	324	177 (55)	84 (26)	0 (0)	23 (7)	0 (0)	40 (12)
Bertelson (1991)	ovarian cancer	716	104 (15)	256 (36)	3 (0)	13 (2)	3 (0)	337 (47)
Unweighted mean proportion §				45	9	9	1	36
Weighted mean proportion §				43	8	6	2	41

Table 1. Group 1 studies: Studies reporting on the enrollment experience of a population cohort to 2 or more randomized controlled trials.

Legend:

† - study reporting enrollment to North American based RCTs;

§ - proportion of subjects for whom a trial was available for their type and stage of disease (ie. denominator is 'sample size' minus 'no protocol available').

Table 2.

Study	Study population	Sample size	Reason for non-participation					Enrolled in trial
			Wrong stage of disease	Ineligible	Physician refused	Subject refused	Other exclusion or unknown	
			Number (percentage)					N (%)
Lee (1980)†	lung cancer	653	434 (66)	142 (22)	34 (5)	2 (1)	0 (0)	41 (6)
CASS (1984)†	heart disease	16626	11838 (71)	2689 (16)	910 (5)	369 (2)	40 (1)	780 (5)
Kronborg (1988)	rectal cancer	1369	508 (37)	341 (25)	2 (0)	9 (1)	24 (2)	485 (35)
Ward (1992)	stomach cancer	1209	93 (8)	310 (26)	?	?	24 (2)	249 (21)
Greil (1993)	affective disorders	6213	2826 (46)	2399 (39)	0 (0)	393 (6)	217 (3)	378 (6)
Unweighted mean proportion §				58	7	4	2	29
Weighted mean proportion §				60	10	8	3	19

Table 2. Group 2 studies: Studies reporting the enrollment experience of a population cohort to a single randomized controlled trial.

Legend:

† - study reporting enrollment to North American based RCT;

§ - proportion of subjects for whom a trial was available for their type and stage of disease (ie. denominator is 'sample size' minus 'wrong stage of disease').

Table 3.

Study	Study population	Sample size	Reason for non-participation				Enrolled in trial
			Ineligible	Physician refused	Patient refused	Other exclusion or unknown	
			Number (percentage)				N (%)
Lucas (1984)	breast cancer	57	41 (72)	5 (9)	8 (14)	3 (5)	0 (0)
Martin (1984)†	cancer patients	2687	1738 (65)	284 (11)	228 (8)	0 (0)	437 (16)
Winger (1989)†	anaplastic glioma	197	119 (60)	0 (0)	23 (12)	0 (0)	55 (28)
Muller (1990)	acute MI	49556	33230 (67)	?	?	?	?
Fentiman (1991)	breast DCIS	216	129 (60)	4 (2)	6 (3)	0 (0)	77 (35)
Hjorth (1992)	multiple myeloma	255	60 (24)	0 (0)	15 (6)	0 (0)	180 (70)
Morris (1993)	acute stroke	200*	87 (44)	?	?	?	?
			191 (96)	?	?	?	?
Unweighted mean proportion §			56	4	9	1	30
Weighted mean proportion §			61	9	8	0	22

Table 3 Group 3 studies: Studies reporting the enrollment of a cohort for whom a trial was available for their type and stage of disease in a single randomized controlled trial.

Legend:

† - study reporting enrollment to North American based RCT;

* - two-hundred patients in an acute stroke unit were assessed for their eligibility for two RCTs that had not yet begun to enroll patients;

§ - proportion of subjects for whom a trial was available for their type and stage of disease (ie. denominator is 'sample size').

Figure 1.

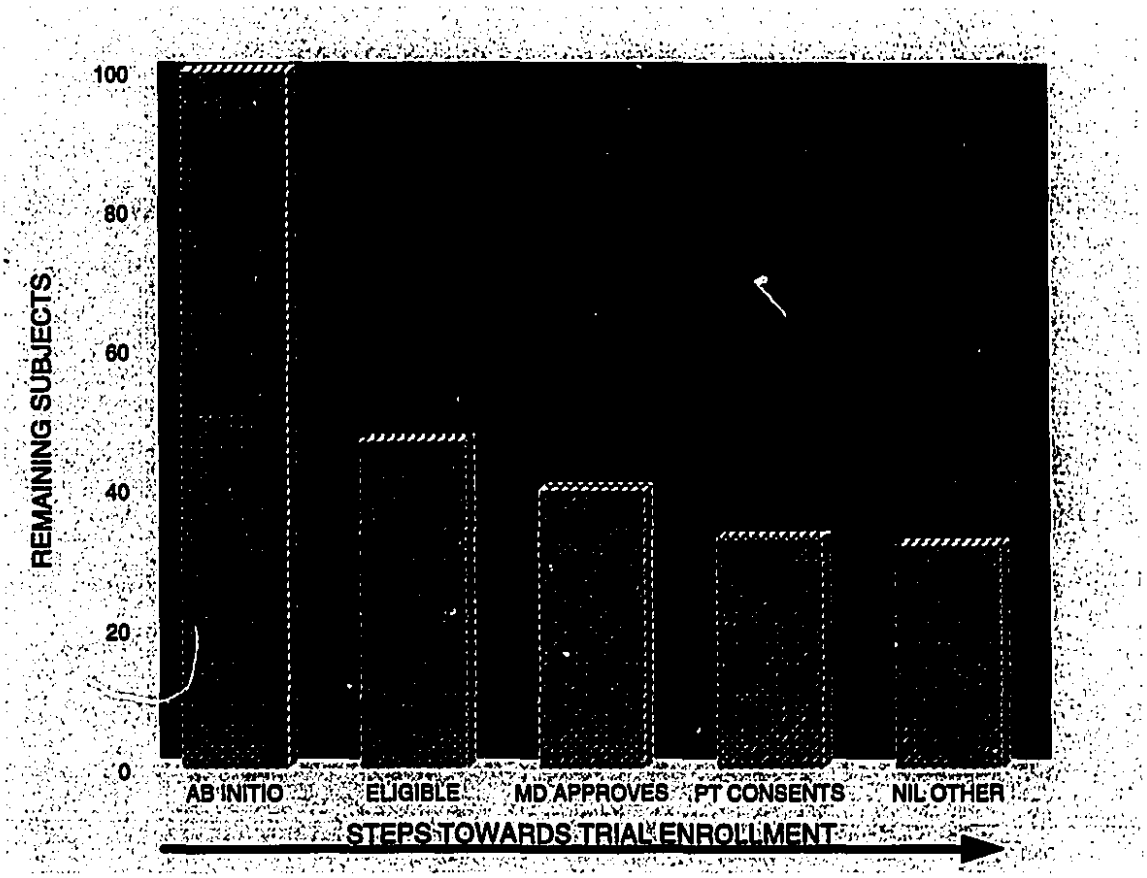
100
80
60
40
20
0
REMAINING SUBJECTS
AB INITIO
ELIGIBLE
MD APPROVES
PT CONSENTS
NIL OTHER
STEPS TOWARDS TRIAL ENROLLMENT

Figure 1. The typical fate of 100 potential subjects for a randomized controlled trial.

References

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

2. Friedman MA, Cain DF. National Cancer Institute sponsored cooperative clinical trials. Cancer 1990; 65: 2376-2382.

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

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

5. 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.

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

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

8. Greco AO, Perry MC. Factors influencing patient accrual to protocol studies. Proceedings of the American Association for Cancer Research and the American Society for Clinical Oncology 1980: C-146.

9. McCusker J, Wax A, Bennett JM. Cancer patient accessions into clinical trials. American Journal of Clinical Oncology 1982; 5: 227-236.

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

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

12. Hunter CP, Frelick RW, Feldman AR, et al.. Results from the Community Clinical Oncology Program Physician's Patient Log. Cancer Treatment Reports 1987; 71: 559-565.

13. 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.

14. Jack WJL, Chetty U, Rodger A. Recruitment to a prospective breast conservation trial: why are so few patients randomized? British Medical Journal 1990; 301: 83-85.

15. Bertelson K. Protocol allocation and exclusion in two Danish randomized trials in ovarian cancer. British Journal of Cancer 1991; 64: 1172-1176.

16. Lee JY, Marks JE, Simpson JR. Recruitment of patients to cooperative group clinical trials. Cancer Clinical Trials 1980; 3: 381-384.

17. CASS Principal Investigators and Their Associates. Coronary Artery Surgery Study (CASS): a randomized trial of coronary artery bypass surgery. Journal of the American College of Cardiology 1984; 3: 114-128.

18. Kronborg O, Fenger C, Bertelson K, et al.. Escape clauses in a multicentre trial: unforseen problems and their potential influence on the general validity of the conclusions. Theoretical Surgery 1988; 2: 157-164.

19. Ward LC, Fielding JWL, Dunn JA, Kelly KA. The selection of cases for randomized trials: a registry survey of concurrent trial and non-trial patients. British Journal of Cancer 1992; 66: 943-950.

20. Greil W, Ludwig-Mayerhofer W, Steller B, et al.. The recruitment process for a multicenter study on the long-term prophylactic treatment of affective disorders. Journal of Affective Disorders 1993; 28: 257-265.

21. Lucas MG, Mitchell A, Lee ECG. Failure to enter patients to randomized study of surgery for breast cancer. Lancet 1984; 2: 921-922.

22. Martin JF, Henderson WG, Zacharski LR, et al.. Accrual of patients into a multihospital cancer clinical trial and its implications on planning future studies. American Journal of Clinical Oncology 1984; 7: 173-182.

23. Winger MJ, MacDonald DR, Schold C, Cairncross JG. Selection bias in clinical trials of anaplastic glioma. Annals of Neurology 1989; 26: 531-534.

24. Muller DWM, Topol EJ. Selection of patients with acute myocardial infarction for thrombolytic therapy. Annals of Internal Medicine 1990; 113: 949-960.

25. Fentiman IS, Julien JP, van Dongen B, van Geel B, Chetty U, Coibion M. Reasons for non-entry of patients with DCIS of the breast into a randomized trial. European Journal of Cancer 1991; 27: 450-452.

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. Morris AD, Grosset DG, Squire IB, Lees KR, Bone I, Reid JL. The experience of an acute stroke unit -- implications for multicentre acute stroke trials. Journal of Neurology, Neurosurgery and Psychiatry 1993; 56: 352-355.

28. 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.

29. Carter SK. Clinical trials in patients with cancer. New England Journal of Medicine 1982; 306: 1105-1106.

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

31. Fuks A, Weijer C, Freedman B, Shapiro S, Skrutkowska S, Riaz A. A study in contrasts: the analysis of eligibility criteria in NSABP breast cancer trials (abstract). Lancet Conference: The Challenge of Breast Cancer. Brugge, April 21,22, 1994.

32. 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.

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

34. Muller DWM, Topol EJ. Selection of patients with acute myocardial infarction for thrombolytic therapy. Annals of Internal Medicine 1990; 113: 949-960.

35. Gurwitz JH, Nananda FC, Avorn J. The exclusion of the elderly and women from clinical trials in acute myocardial infarction. JAMA 1992; 268: 1417-1422.

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

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

38. Yusuf S, Collins R, Peto R et al.. Intravenous and intracoronary fibrinolytic therapy in acute myocardial infarction: overview of results of mortality, reinfarction and side effects from 33 randomized controlled trials. European Heart Journal 1985; 6: 556-585.

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

Bridging section

The second chapter has addressed two of the three empirical questions posed in chapter 1. The first question asked: Is a large proportion of patients with a given disease treated in clinical trials? Although relatively little empirical information exists on this point, it is clear that for the majority of cancers only a small proportion of patients are treated in clinical trials. As a result, we cannot be certain that the results of trials are widely applicable to clinical practice. When the proportion enrolled is small, trial participants may differ for the patient population at large in clinically important ways.

The second question was: Are subjects randomly selected from the patient population for participation in clinical trials? The answer to this question is clearly "no". Protocol factors, physician factors and patient factors have all been cited as being barriers to clinical trial enrollment. The empirical study in chapter 2 attempted to quantify the relative selective force that each of these exerts on trial enrollment. Overall, 32% of subjects for whom a trial was available for their type and stage of disease (and who themselves were available for enrollment) were entered into clinical trials. Eligibility criteria were the largest barriers to trial enrollment, eliminating 53% of

potential subjects. Physician and patient factors were relatively less important, each accounting for roughly 7% of potential subjects (1% were excluded for other reasons).

Since eligibility criteria frequently exclude persons on the basis of clinically important factors -- demographic and prognostic factors, we must be concerned that clinical trial participants may differ in important ways from the population in clinical practice. As such, we must be concerned that the results of trials on such a select population may not be widely generalizable.

The third chapter addresses the issue of the comparability of trial participants and non-participants directly. If trial participants do indeed differ in important ways from trial non-participants, the generalizability of trial results may not be widely applicable. The results thus far would suggest that a solution to this problem will include changes to criteria for trial eligibility.

Chapter 3:

Are randomized controlled trial participants comparable to non-participants?
-- A review of the empirical literature.

Abstract

Introduction. The importance of a randomized controlled trial (RCT) is, in part, a function of how widely applicable its results are to clinical practice. The literature gives us reason to be concerned that the population treated in clinical trials may differ in clinically important ways from the patient population in general: first, only a small proportion of patients are treated in trials and, second, barriers to trial enrollment, such as eligibility criteria, exist. We set out to systematically examine empirical studies that compared trial participants and non-participants.

Methods. Nine variables of interest -- demographic, prognostic and outcome variables -- were defined a priori for this study. Empirical studies of trial participants and non-participants that reported on at least one variable of interest were retrieved from the Medline (1966-1994) and Cancerlit (1983-1994) databases.

Results. In all, 19 studies were retrieved from the literature, the majority of which were cancer clinical trials. RCT participants were significantly younger than non-participants and may have had better performance status scores. These differences were not seen in studies in which

the comparison group included only eligible non-participants, suggesting that eligibility criteria were the cause. Participants also survived longer than non-participants. This significant survival difference remained in six of the seven studies that adjusted survival for important covariates.

Discussion and Conclusion. Important differences exist between patients studied in cancer clinical trials and the patient population at large. The generalizability of RCTs to clinical practice is therefore suspect. Attempts to remedy this situation will need to address the role that criteria for trial eligibility have in unduly narrowing the study population in trials.

Introduction

Any writer who is worth her salt knows that one has to remember for whom one is writing; a text that is meaningful for one audience may not be meaningful, or for that matter interesting, for another. As with skilful writing, so too a randomized controlled trial has an audience who must be considered at the time a trial is designed. Since the importance of a RCT is measured by its impact on clinical practice, trial designers must carefully consider what population of patients a trial outcome is intended, at least ideally, to affect eventually. Decisions made at the time of a trial's design will have a large impact on the scope of the applicability of the trial's results. If the trial includes therapeutic interventions or imaging procedures that are only available in a few technologically-advanced tertiary care centres, the generalizability of the trial's results may be limited. Criteria for trial eligibility will also influence the applicability of a trial. Restricting the study population to only those at very high risk of an outcome -- such as mortality or myocardial infarction, while potentially advantageous from the point of view of reducing the required sample size for a trial, will limit the generalizability of the trial's results. Thus, the intended

impact of a RCT ought to guide decisions made at the time of trial design.

The related issues of trial generalizability and selection of subjects for trials are much disputed in the literature on research design. Following Schwartz and Lellouch's[1] distinction between approaches to RCT design, we might characterize this debate as being between those arguing for what Feinstein[2] refers to as fastidious (scientific, narrow) trials and those calling for pragmatic (clinically relevant, broad) trials.

Proponents of fastidious trials argue that it is important to reduce variability in a trial in order to attempt to isolate the effect of the treatment on the outcome measure and hence, relatively homogenous subgroups of patients ought to be studied in RCTs.[3] Once a treatment has been demonstrated to be effective, the effect may be assumed to have a biological basis, or "biological effect", which, barring evidence to the contrary, may be presumed to be broadly generalizable.[4] The results of RCTs may, thus, be applied to population groups not included in a trial, unless evidence (or strong theoretical basis) exists that a given population group may respond differently to the treatment.

Supporters of pragmatic trials counter that homogeneity of a study population is an unachievable ideal when dealing

with human subjects.[5] Patient-to-patient variability will remain the largest source of variability in most RCTs.[5] This fact is mirrored in clinical practice: while a given disease, such as metastatic breast cancer, has a definable median survival, the prognosis of an individual patient (and her response to treatment) is difficult to predict accurately. The answer to this problem, the clinical-trial pragmatists argue, is to accrue large numbers of patients to trials and this suggests relaxing criteria for trial eligibility.[6] The results of such a trial will be based on a broad-patient population, one that is likely to be reflective of clinical reality.[7] The results of such a trial are, therefore, broadly generalizable.

Recently, in the United States, politicians and patient activist groups have entered the fray. Responding to the alleged under representation of women[8,9] and minorities[10] in clinical trials, recent regulatory changes, including the "NIH Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research",[11] require the inclusion of representative numbers of individuals from these groups in government funded biomedical research.[12,13] In addition, the NIH Guidelines require that, in the absence of strong a priori evidence one way or the other, the analysis of a trial's results include subgroup analyses to examine

possible differential treatment effects in racial/ethnic and gender subgroups.[11] The subgroup analyses, though, are primarily intended as hypothesis generating and, as such, "the trial will not be required to provide high statistical power for each subgroup".[11] Only in the case in which strong a priori evidence exists that a given group will respond differently to a treatment is the trial required to address this as a primary question in the study, that is, to ensure a comparison with sufficient power for that purpose.[11]

These changes in U.S. Government policy were motivated by a number of factors, including concerns that the exclusion of women and racial/ethic minorities from clinical trials may represent discrimination and, furthermore, that these groups may be deprived of any benefits associated with trial participation. Clearly, though, concern over the generalizability of RCTs that fail to include (adequate numbers of) women and racial/ethnic minorities also motivated these changes:

> 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 behavioural 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.[11]

In essence, the NIH's position is that only trials that contain adequate numbers of women and minorities can be considered generalizable to these populations (even in the absence of a priori data to indicate that a differential effect in these groups is likely). As such, the NIH's stand represents a clear rejection of the "biological effect" model of inference to populations of patients more heterogeneous than those actually studied.

Subjects who end up on RCTs may, of course, differ from clinical populations in ways other than gender and race/ethnicity. In a review of factors that influence accrual to RCTs, Gotay describes a range of characteristics that differentiate cancer trial participants from non-participants.[14] According to Gotay, subjects who enter trials tend to be younger, weigh more, have better performance status and have higher socioeconomic status than non-participants. She points out that these differences are caused by obstacles to trial accrual, including criteria for trial eligibility, physician selection and patient factors.[14] A recent review of empirical studies on enrollment in RCTs indicates that criteria for trial eligibility represent the largest selective pressure on

trial enrollment.[15] Of subjects for whom a trial was available for their type and stage of disease, 32% were enrolled in a RCT, 53% were ineligible for trial participation, 7% were excluded due to physician refusal, 7% refused to participate, and 1% were excluded for other reasons.[15]

A review of the literature was undertaken to attempt to address the following questions: Is there evidence that trial participants differ from non-participants? If such differences exist, are they, as suggested above, the result of criteria for trial eligibility? Can an approach similar to that outlined in the NIH Guidelines be applied to rectify the situation?

Methods

Publications of empirical studies reporting on differences between randomized controlled trial participants and non-participants were retrieved through searches of the Medline and Cancerlit databases (Medline: 1966 through October, 1994; Cancerlit: 1983 through September, 1994). The following search terms were used: "clinical trials", "prognosis", "patient participation", "eligibility determination" and "registries". The results of these searches were supplemented with articles referenced in the

bibliographies of relevant articles and other published sources. Empirical studies were included in this review if they reported results of at least one variable of interest (listed infra) for both subjects enrolled in a randomized controlled trial and trial non-participants. Reports of randomized controlled trials for any disease site reported in the English literature were included in this review.

Since eligibility criteria are thought to be a major cause of differences between trial participants and non-participants, comparisons between RCT participants and eligible non-participants are reported separately from comparisons of trial participants and unselected (eligible and ineligible) non-participants. Empirical studies that reported both comparisons for a single RCT are reported twice, once in each category. If criteria for trial eligibility are indeed the major cause of differences between trial participants and non-participants, differences between participants and non-participants should be restricted to the comparison involving unselected non-participants.

The following basic information was recorded for each empirical study: first author, year of publication, study disease, and sample size. The sample size was defined as the number of patients in the RCT plus the number of non-

participants included in the comparison. Thus, a study reporting both comparisons of interest (RCT participants versus eligible non-participants and RCT participants versus unselected non-participants) will have two different sample sizes.

The variables of interest selected for this study were chosen a priori on the basis that they (1) were clinically important, (2) included demographic, prognostic and outcome variables, and (3) were likely to be reported in the published report of a RCT. The following variables were selected: age, gender ratio (male/female), race ratio (Caucasian/non-Caucasian), socioeconomic status, extent of disease, performance status, crude survival, and survival adjusted for covariates. Information on each of these variables was recorded, if available, from each empirical study included. A difference, either "in favour" of RCT participants or trial non-participants, for each of the variables was defined as a statistically significant difference if reported as such in the journal article. In some cases no statistical test was reported in the publication, but sufficient information was present for an unpaired t-test or a chi-square test to be done (as appropriate). In the few cases in which neither a statistical test was reported nor could one be done based on

reported figures, a difference was reported as being present if the authors concluded that the observed difference was clinically important. If none of these conditions obtained, it was concluded that no difference existed between RCT participants and non-participants for that variable. The direction of the difference (if present) was recorded, but not the magnitude of the difference. The studies included in this review cover a heterogeneous group of diseases and, therefore, the magnitude of difference, e.g. months of added survival, was not likely to be meaningfully comparable across studies.

The proportion of studies in which a difference was found (the denominator being the total studies that reported information on a given variable) is reported for each of the variables. Only descriptive statistics are reported.

Results

Nineteen empirical studies that satisfied the criteria for eligibility for this study were retrieved from the literature. The majority of the studies involved persons with cancer (16/19 studies), two studies involved patients with cardiovascular disease, and one study included persons with psychiatric disease (tables 1 and 2). Two studies reported both comparisons of interest and, hence, are

reported in each category of study. Nine studies (published between 1979 and 1993) reported comparisons of RCT participants and eligible non-participants (table 1). In total, the nine studies documented the characteristics of 6,620 individuals. A total of twelve studies (published between 1983 and 1992) document comparisons of RCT participants and unselected non-participants (table 2). These twelve studies reported on 71,820 subjects.

Studies that compared trial participants with eligible non-participants are listed in table 1.[16,17,18,19,20,21,22,23,24] No consistent pattern of differences is seen amongst the demographic variables (table 1). Of the seven studies reporting on age, five (5/7) found no difference between the two groups and two (2/7) reported that subjects on trial were younger. Five of the six studies (5/6) reporting on gender found no difference between the two groups, while one (1/6) found that males were more likely to be enrolled in a RCT. Both of the studies (2/2) reporting on racial characteristics found no difference between study participants and non-participants. Of the two studies reporting on socioeconomic status, one (one of two) reported that trial participants had a higher socioeconomic status and the other (one of two) found no difference between the two groups.

Similarly, there is little evidence for a consistent pattern of difference in prognostic or survival variables (table 1). Of the five studies that reported on extent of disease, four (4/5) found no difference between the groups, while one (1/5) reported that trial participants had more extensive disease. The two studies that reported a difference with regard to performance status were evenly split: one (one of two) reported that trial participants had a higher (better) performance status; the other (one of two) reported that non-participants had a higher performance status. Six studies reported on crude (unadjusted) survival: four (4/6) found no difference between the two groups; two reported a survival advantage for RCT participants. Four studies reported figures for survival adjusted for various covariates. Though two studies (2/4) found no difference, two reported that RCT participants survived longer than non-participants.

Overall then, no consistent pattern of differences was observed in comparisons of RCT participants and eligible non-participants. At least with respect to the variables surveyed, RCT participants are comparable to eligible non-participants. If substantial differences were introduced by factors other than criteria for trial eligibility, for example, physician selection and patient refusal of informed

consent, we would have expected to find differences in this comparison. We did not. It is unlikely, therefore, that these other factors have a substantial impact on shaping the characteristics of trial participants.

Studies that compared RCT participants with unselected (eligible and ineligible) non-participants are listed in table 2.[25,26,27,28,29,30,31,21,32,33,23,34] Substantial differences were noted amongst the demographic variables reported on (table 2). Of the nine studies that reported the age of subjects, seven (7/9) reported that RCT participants were significantly younger than non-participants; two (2/9) reported no difference. One of the two studies that found no age difference between the two groups was a study of childhood leukemia that used an age cutoff of 15 years.[25] It is of note that all of the studies reporting on gender (6/6) and race (2/2) found no difference between the groups. No study reported on socioeconomic status.

Unfortunately, too few studies reported on prognostic variables for there to be a clear indication of a pattern (table 2). All three studies (3/3) that looked at extent of disease found no difference between trial participants and unselected non-participants. Both (2/2) studies that examined performance status found that trial participants had higher scores and, hence, a better performance status.

The data on outcome variables indicate that trial participants survived longer than non-participants (table 2). All nine studies (9/9) reporting crude survival found that RCT participants had higher survival rates than non-participants. All five studies (5/5) that reported on survival adjusted for covariates found a survival advantage for trial participants.

Substantial differences, then, were found between RCT participants and unselected non-participants. Trial participants tended to be younger and survive longer than non-participants. Also, they may have had a better performance status than non-participants (although only two studies reported this). Thus, RCT participants are not comparable, in important ways, with unselected non-participants. As these differences were not seen in comparisons of RCT participants and eligible non-participants, it is likely that these discrepancies are the result of criteria for trial eligibility.

The issue of survival of RCT participants is worthy of further attention. What might account for the survival advantage observed for RCT participants? Table 3 lists the results of the seven studies that both reported a crude survival advantage for trial participants and then adjusted survival for various covariates. The table lists the

method(s) used in the adjustment, the covariates adjusted for and the resulting adjusted survival comparison. If trial participants differ from non-participants in terms of age and other prognostic variables, we might expect them to have a better survival than non-participants on this basis alone. Although some of the studies adjusted for age (5 studies), extent of disease (4 studies) and performance status (1 study), all but one of the studies (6/7) continued to show a survival advantage for trial participants (all of the six "positive" studies adjusted for age or extent of disease, two adjusted for both). A number of possibilities emerge based on this. First, it is possible that the survival difference is indeed due to prognostic differences between the two groups and that we are just not very good at correcting for these differences. Noting from table 2 that many variables were not examined, the answer may lie in filling in the blanks in the table. A second possibility, though, is that RCT participation may, in and of itself, confer a survival advantage on participants. RCTs offer more rigorous protocols for treatment administration and follow-up. It is possible that these factors may translate into a survival advantage for participants.

Finally, in keeping with the theme of this paper, it should be emphasized that given the small number of studies

included in this review, the generalizability of these findings is uncertain. As noted above, the majority of studies involved cancer patients and, thus, the findings are most immediately applicable to cancer trials. The small number of non-cancer studies make the validity of an inference beyond the cancer trial setting dubious.

Discussion

Given the political and regulatory attention recently devoted to the exclusion of women and racial/ethnic minorities from clinical trials, it is surprising that none of the studies in our review found significant differences with respect to gender or race. Indeed, we did find that an important demographic difference exists between trial participants and non-participants: patients who end up on trials tend to be younger. Can we make sense of this finding in light of the literature on the exclusion of women from trials?

While there is evidence that women have been excluded from phase III clinical trials for some categories of disease, evidence for a general phenomenon is lacking. A review by the General Accounting Office (GAO) of New Drug Applications to the Food and Drug Administration (FDA) (1988-1991), showed that women were indeed under represented

in phase II and phase III trials of new cardiovascular agents.[8] There was no evidence, however, of the exclusion of women for any other class of agents included in the study. Bird reviewed clinical research published in the Journal of the American Medical Association in 1990 and 1992 and found that, in studies of non-gender-specific diseases, women were under represented 2.7 times as often as men.[9] These findings were largely explained, though, by the substantial under representation of women in cardiovascular trials.[9]

Is there a connection between the exclusion of women from cardiovascular trials and the exclusion of older patients from trials? It seems that there may be one. Gurwitz et al reviewed 214 RCTs of the treatment of myocardial infarction and found that the proportion of women included in trials was strongly associated with the presence of age-based exclusion criteria in the trial.[35] RCTs with age-based exclusion criteria had a significantly lower proportion of women enrolled in the study. This is, of course, a reflection of the fact that women get heart disease later in life than men; excluding older persons from cardiovascular trials, therefore, differentially excludes women from trials. So it seems that gender differences in cardiovascular trials may, at least in part, be explained by age-based exclusions rather than gender-based exclusions per

se. The results of our study suggest that under representation of the elderly may be more wide-spread than gender under representation. If true, the focus of the NIH Guidelines seems misplaced.

The exclusion (or under representation) of groups of patients from RCTs can have serious consequences. The exclusion of the elderly, particularly from cancer trials, has lead to a serious lack of information regarding the proper treatment of older persons with cancer, the population most heavily burdened with the disease.[36] A number of studies have demonstrated that older persons with cancer tend to be under-treated.[37,38] This phenomenon may, in part, be due to the paucity of clinical trials addressing the treatment of cancer in the elderly.[39,40]

Older persons with cancer may differ from younger persons with cancer in clinically important ways. Older persons have long been thought to be more susceptible to side effects from chemotherapy and other cancer treatments.[41] It is certainly true that some chemotherapy agents induce more frequent or more severe side effects (or both) in older persons.[42] Also, important biological differences in tumours may be related to age. For example, cancers of the colorectum, lung, prostate and bladder have been reported to be of a less differentiated histological

grade in older patients.[43] The net effect of these differences is that response to cancer treatments in the elderly is diverse -- some cancer types showing better responses (e.g., colon cancer),[44] some similar responses (e.g., lung cancer),[45] and some worse responses (e.g., acute leukemias, lymphomas, and Hodgkin's disease)[46,47,48] than cancers in younger patients. These differences are not a priori predictable from studies done solely on younger cancer patients.

The heterogeneity in response to treatment in the elderly highlights the basic fallacy with the "biological effect" model: even if biological responses (of tumours, say) are generalizable from narrow study populations, other host factors may differ greatly in subgroups of the target population leading to differing net responses to treatment. (An aggressive treatment regimen that eradicates the cancer but kills the patient is of little benefit.) Study populations will need to be broad-based to examine these issues adequately.

But how many subjects from a given subgroup need to be included in a trial in order to provide a sound basis for the treatment's generalizability to clinical practice? I take it as uncontroversial that if no subjects from a group defined by some factor of potential biological (or clinical)

importance are included in a trial, no valid inference can be made regarding the treatment's efficacy in that group.[49] But, is the inclusion of a representative proportion of individuals from a given group in a trial a sufficient basis to conclude that the treatment is proven effective in (i.e., can be generalized to) that group? Or, must we prove the effectiveness of a treatment for each group by means of a formal demonstration of a statistically and clinically significant difference in favour of the study treatment?

A conservative attempt at a solution to this problem might begin by considering the plausibility of the following premise: A trial large enough to allow for a separate and sufficiently powerful comparison for each group of biological or clinical interest is to be preferred over a smaller trial designed around the average (overall) effect. Clearly, depending on the number of groups of biological or clinical interest, the former trial will be much larger indeed than the latter. Consider the example given by Yusuf et al.[50]: Suppose one is planning a RCT to examine the effect of ß-blockers on mortality post-myocardial infarction. The expected mortality in the control group is approximately 10% and the expected mortality in the experimental arm is 7.5% (a relative mortality reduction of 25%). If power is set at 90% and significance level at 0.01,

then just under 4000 subjects will be required in each of the two study arms, a total of about 8000 subjects. Suppose now that we wish to examine the effect of the treatment separately in two groups, anterior infarctions and inferior infarctions. Holding all else constant, the trial will now require 16000 subjects. Of course, there may be many other groups of interest -- women versus men, older versus younger, class of cardiovascular disability (4 groups) -- and the sample size of the trial will increase proportional to the product (2x2x4) of the number of categories of interest. Clearly, the feasibility of such an approach to trials is questionable -- a single trial may consume all the resources that a funding agency has available.

But there is an even more vexing problem with this approach. If we assume that the treatment effect is constant across the groups of interest (and that accrual to each of the groups is uniform), then an analysis for average effect will reach statistical significance long before the comparisons in the individual groups. This will leave the investigators in the uncomfortable position of "knowing" that <u>on average</u> the experimental treatment is superior to the control treatment, and yet being forced to continue the trial to allow this effect to be proven in each of the subgroups. As such, continuing the trial in this

circumstance seems to violate the Declaration of Helsinki, the most influential international statement on the ethics of research, which requires that: "In any medical study, every patient -- including those of a control group, if any -- should be assured of the best proven diagnostic and therapeutic method".[51]

Underlying this dilemma is a basic question of medical epistemology -- when has a treatment been proven effective?; when ought/must we stop a trial? -- that gets at the heart of the ethical permissibility of clinical research. One approach to this problem is to examine the conditions that must exist ab initio for a trial to be ethical; if and when, during the conduct of the trial, these conditions no longer obtain, the trial must be stopped.

There is a consensus that at the beginning of a RCT comparing two or more treatments an honest null hypothesis must exist.[52] In other words, uncertainty must exist as to the relative merits of the treatments being tested in the trial.[53] Some authors have argued that this means the treatments in a trial must be precisely balanced -- referred to as "theoretical equipoise" -- that is, no empirical grounding for a preference for one treatment over another in a trial can exist.[54] As Freedman has correctly pointed out, this understanding of equipoise is all too fragile: the fate

of a single patient in a RCT could throw the balance in favour of one treatment or the other, thus, requiring that the trial be stopped.[55] Freedman has persuasively argued for a different understanding of equipoise termed "clinical equipoise". Clinical equipoise exists when there is "an honest, professional disagreement among expert clinicians about the preferred treatment".[55] A trial is permissible if the following conditions obtain:

> [A]t the start of the trial, there must be a state of clinical equipoise regarding the merits of the regimens to be tested, and 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.[55]

If the yardstick of a successful (and ethical) RCT is its ability to settle the dispute with regard to treatment preference among expert clinicians, then the need for an overall demonstration of the superiority of one agent over another versus the demonstration of superiority in a set of clinical relevant subgroups will depend on the skepticism in the expert community with regard to the universal applicability of the treatments to a patient population. Since medical practice is evidence based, this skepticism will largely be a function of the existence of evidence (or substantial theoretical considerations) that one or more of

the treatments is likely to have substantially different effects in one or more subgroups. In the absence of strong evidence for such a differential effect, I contend, the demonstration of the superiority of one treatment in the aggregate is likely to be sufficient to resolve the disagreement among expert practitioners. Since, at this point, clinical equipoise is disturbed, the conditions for the ethical permissibility of the trial have been altered and, hence, the trial must be stopped.[56] One treatment is now (at least in potentio) held by expert practitioners to be superior and therefore continuing to enroll subjects to the other treatment arms would be unacceptable since they would be deprived of the "best proven...therapeutic methods".

If, however, strong evidence exists that a subgroup of the patient population may respond substantially differently -- perhaps the group is thought to be more likely to suffer serious adverse effects or, for some reason, is thought less likely to benefit from the treatment -- from the rest of the population, the RCT ought to be designed to be sufficiently powerful to convincingly answer the question for each of the groups in question. The demonstration of an advantage for one treatment in the aggregate is, in this case, not likely to sway practice; thus, the trial ought to be continued

until the question is answered for these groups. Since, at this point in the trial (i.e., when only proof for an average effect exists), clinical equipoise has not been disturbed, and, by definition, no patient is disadvantaged in terms of their medical treatment by the trial's continuation. Indeed, since it is a characteristic of an ethical trial that the "results of a successful trial should be convincing enough to resolve the dispute", it would be unethical to stop the trial prematurely.

What is the proper place for subgroup analyses in the first of these two scenarios, i.e., when no strong evidence exists for a different effect? A number of excellent reviews of the use and interpretation of subgroup analyses in those trials designed around the investigation of an aggregate effect have recently appeared.[50,57,58] While the reader interested in an in depth treatment of the subject may consult these sources, several brief points ought to be made here. The first is that the average effect across subgroups is the most reliable indicator of the effect of a treatment. Thus, if the overall result of the trial is negative, subgroup analyses are to be discouraged. Conversely, if the overall result of a trial is positive, i.e. the experimental treatment is shown (overall) to be superior to the standard treatment, the absence of a statistically significant effect

in some subgroups should not be interpreted as a lack of efficacy in those subgroups since the comparisons are likely to be substantially under-powered. Returning to our example of the ß-blocker trial with 8000 subjects enrolled, if the true effect of the treatment is a 25% mortality reduction and this is true for the subgroups of anterior and inferior myocardial infarctions, an analysis done separately for each of the two subgroups will only have a power of 56%. Thus, assuming that the subgroup analyses are independent of each other, there is only a 31% chance that both subgroup analyses will conclude that the treatment is effective. Subgroup analyses should be understood as being useful for generating hypotheses for future study.

Much of what we have concluded regarding the interpretation of treatment effects for subgroups of the patient population is mirrored in the "NIH Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research". And as such, an approach like that outlined in the NIH Guidelines could be used to approach clinical research in general. In summary the approach would involve the following:

- Phase III clinical trials ought to include subjects who are as representative of the population in clinical practice as possible. As eligibility criteria seem to be the major

cause of a lack of representativeness, the set of criteria chosen should be minimized -- the fewer criteria, the better. The explicit justification of each criterion will help to emphasize its necessity.

- When strong evidence for a different effect of a treatment in a clinically relevant subgroup of patients -- such as women, older persons or Afro-Americans -- sufficient subjects ought to be accrued to enable the question of the treatment's efficacy to be answered separately for that group. Typically, this will require a that a trial be substantially larger. It must however be recalled that in these circumstances, a smaller trial is unlikely to change clinical practice for all groups of patients.
- If there is no strong evidence <u>a priori</u> for a different effect in any clinically relevant patient group -- and I expect this circumstance will be the rule -- the use of subgroup analyses in trials studying the aggregate effect is still to be encouraged. These analyses, though, should be understood as hypothesis generating and, as discussed above, the temptation to treat their results as conclusive should be avoided.

I would be remiss in my duties if I did not comment on the survival differences observed between trial participants and unselected non-participants. In the cancer literature,

it has long been suggested, but never definitively proven, that subjects treated in RCTs do better. Stiller, both in an editorial in the British Medical Journal[59] and in a more thorough review elsewhere,[60] has suggested that the treatments given to cancer patients on trial and the fact that care is usually given in a tertiary centre both contribute to improved survival. Our results clearly support the assertion that trial participants enjoy better outcomes that unselected non-participants. Whether though this is due to trial participation per se or is a mere reflection of the selective effect of other factors such as eligibility criteria, however, we cannot say. On the one hand, the fact that only two of six studies that compared trial participants with eligible non-participants showed this difference suggests that the survival advantage may be in large part due to a selective effect. On the other hand, the failure of adjustment for well-known prognostic variables to correct for observed survival differences argues for an advantage above and beyond a mere selective effect. The issue has not been solved by this study and awaits resolution by a more comprehensive approach.

Table 1.

Study	Study population	Sample size	Demographic variables				Prognostic variables		Outcome variables	
			age	gender ratio	race ratio	socioeconomic status	extent of disease	performance status	crude survival	survival adjusted for covariates
Lennox (1979)	childhood nephroblastoma	202	0	0	0	0	0	0	+	+
McCusker (1982)	lung cancer	174	=	=	=	+	=	-	=	0
CASS (1984)	coronary artery disease	2 095	=	=	=	=	=	0	=	=
Antman (1985)	sarcoma	90	=	=	0	0	+	0	0	+
Smith (1988)	coronary artery disease	1 484	-	+	0	0	0	0	0	0
Winger (1989)	anaplastic glioma	78	=	0	0	0	0	+	=	0
Bertelson (1991)	ovarian cancer	481	0	0	0	0	=	0	+	=
Ward (1992)	stomach cancer	710	-	=	0	0	=	0	=	0
Greil (1993)	affective disorders	1 306	=	=	0	0	0	0	0	0

Table 1. Summary of studies comparing RCT participants and eligible non-participants.

Legend:

+ - (compared to non-participants) RCT participants were reported to be older, more likely to be male, more likely to be white, of higher socioeconomic status, have more extensive disease, have a higher performance status, or be more likely to survive;

- - (compared to non-participants) RCT participants were reported to be younger, more likely to be female, more likely to be non-white, of lower socioeconomic status, have less extensive disease, have a lower performance status, or be less likely to survive;

= - no difference between the two groups;

0 - information not reported on in article.

Table 2.

Study	Study population	Sample size	Demographic variables				Prognostic variables		Outcome variables	
			age	gender ratio	race ratio	socioeconomic status	extent of disease	performance status	crude survival	survival adjusted for covariates
Meadows (1983)	acute lymphocytic leukemia	327	=	=	=	0	=	0	+	+
Martin (1984)	various cancers	2 687	-	=	=	0	0	+	0	0
Boros (1985)	acute nonlymphocytic leukemia	130	-	0	0	0	0	0	+	+
Quoix (1986)	small-cell lung cancer	215	-	=	0	0	=	0	+*	0
Hunter (1987)	various cancers	17 773	-	0	0	0	0	0	0	0
Goodwin (1988)	various cancers	42 724	-	0	0	0	0	0	0	0
Michaelis (1988)	childhood osteosarcoma	210	0	0	0	0	0	0	+	0
Winger (1989)	anaplastic glioma	197	-	0	0	0	0	+	+	0
Karjalainen (1989)	multiple myeloma	1 978	=	=	0	0	0	0	+	+
Stiller (1989)	acute lymphoblastic leukemia	4 070	0	=	0	0	0	0	+	+
Ward (1992)	stomach cancer	1 209	0	0	0	0	0	0	+	0
Hjorth (1992)	multiple myeloma	300	-	=	0	0	=	0	+	+*

Table 2. Summary of studies comparing RCT participants and unselected (ie. both eligible and ineligible) non-participants.

Legend:

+ - (compared to non-participants) RCT participants were reported to be older, more likely to be male, more likely to be white, of higher socioeconomic status, have more extensive disease, have a higher performance status, or be more likely to survive;

- - (compared to non-participants) RCT participants were reported to be younger, more likely to be female, more likely to be non-white, of lower socioeconomic status, have less extensive disease, have a lower performance status, or be less likely to survive;

= - no difference between the two groups;

0 - information not reported on in article;

* - for one subgroup only.

Table 3.

Study	Controls	Analytic method(s)				Covariates					Adjusted survival
		design	stratification	logistic regression	proportional hazards	age	gender	stage	treatment	other	
Lennox (1979)	eligible		✔			✔		✔			+
Meadows (1983)	unselected		✔	✔		✔	✔	✔		treatment centre	+
Boros (1985)	unselected	✔			✔			✔	✔	platelet count, LDH, perf. status, antibiotics, preleukemia, fever	+
Karjalainen (1989)	unselected				✔	✔	✔				+
Stiller (1989)	unselected		✔			✔				period of diagnosis	+
Bertelson (1991)	eligible		✔					✔	✔		=
Hjorth (1992)	unselected		✔			✔			✔		+*

Table 3. Studies showing a crude survival advantage for trial participants that also adjusted survival data for covariates: method(s) used and covariates controlled for.

Legend:

Controls indicate whether the study used eligible non-participants ("eligible") or unselected non-participants ("unselected") as the control group. (Outcome data for studies with eligible non-participants as controls are found in table 1; outcome data for studies with unselected controls are found in table 2). Analytic methods include: design, stratified analysis, logistic regression model, and Cox proportional hazards model. Covariates include: age, gender, stage or extent of disease ("stage"), treatment received ("treatment"), and other covariates. Adjusted survival: + - survival advantage for trial participants; - - survival advantage for trial non-participants; = - no difference between the two groups. * - difference for only one subgroup.

References

1.Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. Journal of Chronic Disease 1967; 20: 637-648.

2.Feinstein AR. An additional basic science for clinical medicine II, the limitations of randomized trials. Annals of Internal Medicine 1983; 99: 544-550.

3. 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.

4. Davis CE. Generalizing from clinical trials. Controlled Clinical Trials 1994; 15: 11-14.

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

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

7. 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.

8. General Accounting Office. Women's health: FDA needs to ensure more study of gender differences in prescription drug testing. Washington, D.C.: Government Printing Office, 1992. (Publication no. GAO/HRD-93-17.)

9. Bird CE. Women's representation as subjects in clinical studies: a pilot study of research published in JAMA in 1990 and 1992. In: Mastroianni AC, Faden R, Federman D (eds.). Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies (Volume 2). Washington, D.C.: National Academy Press, 1994: 151-173.

10. Svensson CK. Representation of American blacks in clinical trials of new drugs. JAMA 1989; 261: 263-265.

11. National Institutes of Health. NIH Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research. Federal Register 1994; 59(59): 14508-14513.

12. Bennett JC. Inclusion of women in clinical trials -- policies for population subgroups. New England Journal of Medicine 1993; 329: 288-292.

13. Merkatz RB, Temple R, Sobel S, et al.. Women in clinical trials of new drugs: a change in Food and Drug Administration policy. New England Journal of Medicine 1993; 329: 292-296.

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

15. Weijer C. Eligibility criteria and other barriers to enrollment in randomized controlled trials. Manuscript under submission.

16. Lennox EL, Stiller CA, Morris Jones PH, Kinnier Wilson LM. Nephroblastoma: treatment during 1970-3 and the effect on survival of inclusion in the first MRC trial. British Medical Journal 1979; 2: 567-569.

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

18. CASS Principal Investigators and Their Associates. Coronary Artery Surgery Study (CASS): A randomized trial of coronary artery bypass surgery. Journal of the American College of Cardiology 1984; 3: 114-128.

19. Antman K, Amata D, Wood W, et al.. Selection bias in clinical trials. Journal of Clinical Oncology 1985; 3: 1142-1147.

20. Smith P, Arnesen H. Non-respondents in a post-myocardial infarction trial: characteristics and reasons for refusal. Acta Medica Scandinavica 1988; 223: 537-542.

21. Winger MJ, Macdonald DR, Schold C, Cairncross JG. Selection bias in clinical trials of anaplastic glioma. Annals of Neurology 1989; 26: 531-534.

22. Bertelson K. Protocol allocation and exclusion in two Danish randomized trials in ovarian cancer. British Journal of Cancer 1991; 64: 1172-1176.

23. Ward LC, Fielding JWL, Dunn JA, Kelly KA. The selection of cases for randomised trials: a registry survey of concurrent trial and non-trial patients. British Journal of Cancer 1992; 66: 943-950.

24. Greil W, Ludwig-Mayerhofer W, Steller B, et al.. The recruitment process for a multi-center study on the long-term prophylactic treatment of affective disorders. Journal of Affective Disorders 1993; 28: 257-265.

25. Meadows AT, Kramer S, Hopson R, Lustbader E, Jarrett P, Evans AE. Survival in childhood acute lymphocytic leukemia. Cancer Investigation 1983; 1: 49-55.

26. Martin JF, Henderson WG, Zacharski LR, et al.. Accrual of patients into a multihospital cancer clinical trial and its implications on planning future studies. American Journal of Clinical Oncology 1984; 7: 173-182.

27. 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.

28. Quoix E, Finkelstein H, Wolkove N, Kreisman H. Treatment of small-cell lung cancer on protocol: potential bias of results. Journal of Clinical Oncology 1986; 4: 1314-1320.

29. Hunter CP, Frelick RW, Feldman AR, et al.. Selection factors in clinical trials: results from the Community Clinical Oncology Program physician's patient log. Cancer Treatment Reports 1987; 71: 559-565.

30. Goodwin JS, Hunt WC, Humble CG, Key CR, Samet JM. Cancer treatment protocols: who gets chosen? Archives of Internal Medicine 1988; 148: 2258-2260.

31. Michaelis J. Osteosarcoma. Lancet 1988; I: 1174.

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

33. Stiller CA, Draper GJ. Treatment centre size, entry to trials, and survival in acute lymphobastic leukemia. Archives of Disease in Childhood 1989; 64: 657-661.

34. 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.

35. Gurwitz JH, Col NF, Avorn J. The exclusion of the elderly and women from clinical trials in acute myocardial infarction. JAMA 1992; 268: 1417-1461.

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

37. Chu J, Diehr P, Feigl P, et al.. The effect of age on the care of women with breast cancer in community hospitals. Journal of Gerontology 1987; 42: 185-190.

38. Allen C, Cox EB, Manton KG, Cohen HJ. Breast cancer in the elderly: current patterns of care. Journal of the American Geriatric Society 1986; 34: 637-642.

39. 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.

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

41. Frei E, Canellos GP. Dose: a critical factor in cancer chemotherapy. American Journal of Medicine 1980; 69: 585-594.

42. Walsh SJ, Begg CB, Carbone PP. Cancer chemotherapy in the elderly. Seminars in Oncology 1989; 16: 66-75.

43. Holmes FF. Clinical evidence for a change in tumor aggressiveness with age. Seminars in Oncology 1989; 16: 34-40.

44. Calabrese CT, Adam YG, Volk H. Geriatric colon cancer. American Journal of Surgery 1973; 125: 181-184.

45. Kirsh MM, Rotman H, Bove E, et al.. Major pulmonary resection for bronchogenic carcinoma in the elderly. Annals of Thoracic Surgery 1976; 22: 369-373.

46. Lipschitz DA, Goldstein S, Reis R, et al.. Cancer in the elderly: basic science and clinical aspects. Annals of

Internal Medicine 1985; 102: 218-228.

47. Holmes FF, Hearne E, Conant M, et al.. Survival in the elderly with acute leukemia. Journal of the American Geriatric Society 1979; 27: 241-243.

48. Peterson BA, Pajak TF, Cooper MR, et al.. Effect of age on therapeutic response and survival in advanced Hodgkin's disease. Cancer Treatment Report 1982; 66: 889-898.

49. Simon R. Patient heterogeneity in clinical trials. Cancer Treatment Reports 1980; 64: 405-410.

50. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991; 266: 93-98.

51. World Medical Association Declaration of Helsinki: Recommendations guiding medical doctors in biomedical research involving human subjects. In: Levine RJ. Ethics and Regulation of Clinical Research. New Haven, Yale University Press, 1988: 427-429.

52. Levine RJ. Ethics and Regulation of Clinical Research. New Haven: Yale University Press, 1988: 187-190.

53. Fried C. Medical Experimentation: Personal Integrity and Social Policy. Amsterdam: North Holland Publishing, 1974.

54. Shaw LW, Chalmers TC. Ethics in cooperative clinical trials. Annals of the New York Academy of Sciences 1970; 169: 487-495.

55. Freedman B. Equipoise and the ethics of clinical research. New England Journal of Medicine 1987; 317: 141-145.

56.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.

57. Bulpitt CJ. Subgroup analysis. Lancet 1988; 2: 31-34.

58. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Annals of Internal Medicine 1992; 116: 78-84.

59. Stiller CA. Survival of patients with cancer: those included in trials do better. British Medical Journal 1989;

299: 1058-1059.

60. Stiller CA. Survival of patients in clinical trials and at specialist centres. In: Williams CJ (ed.). Introducing New Treatments for Cancer: Practical, Ethical and Legal Problems. Chichester: John Wiley and Sons Limited, 1992: 119-136.

Chapter 4

Discussion: Summary of findings and implications for the ethical review of clinical research

Fitting the results of chapters 2 and 3 into the historical-ethical schema

In the first chapter of the thesis I presented both a normative and a historical context for issues regarding the selection of subjects for clinical research. How do the results described in the second and third chapters of this thesis fit into the picture we have presented? Overall, I have attempted in this thesis to characterize the population(s) of persons studied in clinical trials. As such, this work addresses the issue of the applicability, or generalizability, of the results of trials to clinical practice. Rather than a historical or ethical (normative) approach, though, the work in this thesis adopts an empirical approach to the question. How many patients end up treated on clinical trials? What barriers prevent patients from entering trials and which is the greatest? Are persons who are treated in clinical trials similar -- in terms of demographics, prognosis and outcome -- to persons treated in clinical practice? These are all, of course, empirical questions.

The ordering of the empirical questions in this work was based on logical considerations regarding the generalizability of the results of randomized controlled trials. The questions were posed in order of priority such

that an affirmative answer would terminate the need for subsequent questions. The questions, the reasoning behind them, and the results in brief are as follows:

- Is a large proportion of patients with a given disease treated in clinical trials? If all persons with a given disorder, or at least a very substantial proportion of them, are treated in clinical trials, then, ipso facto the results of those trials would be widely generalizable. We found that the proportion of patients who end up on clinical trials is not well characterized for most diseases. Data from Tate et al[1] and from Friedman et al[2] on the enrollment of cancer patients in clinical trials indicate that only a small proportion of cancer patients (1.6 - 3.7%) is treated in clinical trials. This is likely also the case for clinical trials examining most common diseases. Since only a relatively small proportion of patients are treated in clinical trials, we cannot be assured that the subjects in trials do not differ from the general patient population in important ways. Thus, it may be that the results of trials are not broadly applicable to clinical practice.

- Are subjects randomly selected from the patient population for participation in randomized controlled

trials? For common medical conditions, even if only a small proportion of a patient population is treated in a given clinical trial, we could be reasonably certain that the results of that trial were broadly applicable if the participants were randomly selected from the patient population and if the number of participants was reasonably large. It seems that the selection of subjects for participation in randomized controlled trials is anything but random. Gotay reviewed accrual to cancer clinical trials and found that protocol, physician and patient factors presented significant and differential barriers to patient enrollment[3].

Much of the second chapter is devoted to an attempt to characterize the relative selective force that each of these factors exerts. Determining this is important for two reasons. First, protocol factors, that is, criteria for clinical trial eligibility, select patients on the basis of factors -- such as demographic and prognostic criteria -- that are of obvious clinical importance. Thus, evidence that eligibility criteria are the most significant selective force would immediately cause us to suspect that trial participants differed from non-participants. Second, and more importantly, an understanding of which of these factors is the largest barrier to trial enrollment would direct

attempts to remedy the situation. If eligibility criteria are the largest barrier, then criteria could be more closely scrutinized and, if possible, eliminated; if physician factors are the most important, physicians could be targeted for educational or incentive programs; if patient factors proved to be the main factor, patients could be educated about clinical trials or (more radically) consent modifying procedures could be considered.

The literature reviewed in second chapter summarizes the findings presented in twenty published reports describing the enrollment experience of 148,561 potential research subjects. Of subjects for whom a trial was available for their type and stage of disease (and who themselves were available for the trial), 32% were actually enrolled in a randomized controlled trial. Eligibility criteria were the largest barrier to trial enrollment for these potential subjects: 53% were excluded by criteria for trial eligibility. Physician factors and patient factors each accounted for 7% of the subjects excluded (and 1% were not enrolled for other reasons). These findings indicate that there is good reason to believe that patients treated in clinical trials may not be comparable in important ways to the general patient population. Furthermore, whatever differences exist between persons treated in clinical trials

and those treated in clinical practice are likely the result of criteria for trial eligibility.

● Are clinical trial participants comparable to non-participants? The third chapter tackles the issue of the comparability of patients treated in clinical trials with those treated in clinical practice head on. Nine variables of interest -- including demographic, prognostic and outcome variables -- were defined a priori. In total, 19 empirical studies were retrieved from the literature which compared trial participants and non-participants with regard to at least one of the variables of interest. Building on the results of the second chapter which suggested that any differences observed would likely be due to eligibility criteria, the comparisons were divided into two groups: comparisons of trial participants with eligible non-participants; and, comparisons of trial participants with both eligible and ineligible trial non-participants. If criteria for trial eligibility are indeed the cause of any differences then, we would expect to observe differences in only the second group. The study presented in the third chapter was designed to answer two questions: (1) Do subjects who are treated on trial differ from study non-participants with regard to any of the variables of

interest; and , (2) Are these differences, if any, due to criteria for trial eligibility?

The results presented in the third chapter show that substantial differences do indeed exist between clinical trial participants and non-participants. Subjects enrolled in trials are significantly younger than trial non-participants. Subjects enrolled in trials may also have better performance status scores than non-participants, although too few studies reported on this variable to allow this to be concluded with confidence. These differences were only apparent in comparisons of participants with unselected non-participants and, thus, they were likely due to criteria for trial eligibility. Subjects enrolled in clinical trials also tended to survive longer than patients treated in clinical practice. It could not be determined from the results whether this represented a survival advantage conferred by the (presumably) more fastidious treatments offered in clinical trials or the selective effect exerted by eligibility criteria or some other factor.

Thus, the results of this study underscore the importance of the concern expressed in the last few years by ethicists, politicians and patient activists regarding the applicability of the results of clinical research to the heterogeneous patient population at large. Our failure to

find differences with regard to gender or race between trial participants and non-participants is, however, surprising. Confirmation of this finding will have to await further study.

Nonetheless, the empirical literature which has been used to bolster claims of the "ubiquitous" exclusion of women and racial/ethnic minorities from clinical research bears re-examination. Very few empirical studies have addressed the exclusion of racial/ethnic minorities from research (Svensson's study[4] is the only one of which I am aware) and the contention that minorities have been under represented in clinical research remains unproven. There is more empirical evidence to suggest that women have been systematically excluded from certain categories of phase II and III clinical trials, particularly cardiovascular trials.[5,6] As I argue in chapter 3, though, the gender discrepancy observed in cardiovascular trials may be explained by age-based exclusion criteria.

Recall that (in our review of the history of ethical concerns regarding the selection of subjects, Chapter one) the predominant concern in the 1970s -- the equitable distribution of burden -- was based on the belief that research participation was an activity laden with risk. A number of empirical studies (cited in chapter one)

established that this premise was, for the most part, false. So too, I contend, the emphasis placed on the issue of the exclusion of women and racial/ethnic minorities from clinical research, as exemplified by the new NIH Guidelines, is grounded in a false or exaggerated premise. The exclusion of the elderly from clinical trials may be a more widespread and, hence, more important problem.

Implication of these findings for the review of research by Institutional Review Boards

The obligations of the Institutional Review Board in the review of clinical research are detailed in the Department of Health and Human Services Regulations (revised 1991).[7] The portion of the regulations pertaining to the selection of subjects was cited in extenso in the first chapter and, in brief, it charges the IRB to ensure that the "selection of subjects is equitable" (§46.111(a)(3)). The Institutional Review Board Guidebook interprets this to mean that the IRB must ensure that both the burdens and benefits of the research are equitably distributed.[8] Thus, with regard to the former (burden), IRBs must ensure that "vulnerable" populations (such as those incapable of giving consent) are not unnecessarily or inappropriately included in research. If circumstances so require, IRBs may insist

that a "vulnerable" group be excluded from a given study. With regard to the distribution of benefit, investigators and IRBs are cautioned not to "overprotect vulnerable populations so that they are excluded in from participating in research in which they wish to participate".[8] Thus, IRBs may question criteria for clinical trial eligibility that seem to exclude groups of affected individuals who may benefit from participation without justification. In the absence of an acceptable justification, an IRB may make the elimination of the eligibility criterion in question a condition of the study's approval.

The "NIH Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research", described in some detail in chapter one, require applicants for NIH funding to ensure that women and racial/ethnic minorities are adequately represented in clinical research proposals.[9] The Guidelines gives IRBs a role in ensuring that their objectives are fulfilled. When reviewing applications for NIH funding, IRBs must review procedures for subject selection and ensure that the exclusion or inadequate representation of women or racial/ethnic minority groups is carefully and adequately justified. Furthermore, IRBs have an important educational role to play in the success of the Guidelines. Educating investigators as to the importance of

research with broadly applicable results need not, of course, only be directed at investigators who are applying for NIH funding. So too, institutional guidelines and procedures for enrolling and retaining women and racial/ethnic minorities in research studies need not be restricted to NIH funded research.

A strict interpretation of the new NIH Guidelines does, however, limit the IRB's role in questioning eligibility criteria that limit the generalizability of the study to applications for NIH funding. But, according to the NIH Guidelines, can IRBs only question criteria for trial eligibility that explicitly deal with women or racial/ethnic minorities? Since the stated purpose of the Guidelines is to ensure (in the case of phase III trials) "sufficient and appropriate entry of gender and racial/ethnic subgroups" -- i.e., adequate representation (rather than mere access), it may be argued that IRBs can also direct their attention to eligibility criteria that indirectly hinder the enrollment of women or minorities to trials.[9] Thus, based on our research findings, an IRB may question a criterion in a cardiovascular disease trial that excludes, for example, persons over the age of seventy years <u>on the basis that it will disproportionately exclude women</u> from trial participation. It will, of course, be the IRB's obligation

to establish the connection between the criterion in question and gender or minority enrollment.

But what of research that does not fall under the NIH Guidelines? What is the place for a critical assessment by the IRB of the generalizability of a study in the Department of Health and Human Services Regulations? Let me highlight one possible approach to this difficult problem with an example.

Recall that in the discussion section of chapter three we presented a hypothetical randomized controlled trial which involved testing a new drug versus placebo in the setting of post-myocardial infarction.[10] The endpoint of the study was mortality and the sample size was calculated on the basis of an expected 10% mortality in the control group, 7.5% mortality in the experimental group, power of 90%, and probability of type I error of 1%. Given these assumptions, the required sample size is (approximately) 4000 persons per group or about 8000 in total. To bring out the issues at play here, I would like to add the following "complications" to this scenario:

- The investigator has non-NIH funding.
- The investigator, mindful of the political tenor of the times, plans to enroll only women in the study. She outlines

plans for actively accruing women to the study and projected accrual goals that seem realistic.

- Strong evidence exists that men and women may respond quite differently to the proposed treatment regimen. (For the purposes of the example, allow me to remain somewhat vague about this). The strength of evidence is such that the committee feels, after consulting with a number of expert practitioners, that, were the issue of gender difference addressed in the study, it would have to be addressed in the primary analysis of the study.
- Finally, with regard to the study itself, let us assume that all of the interventions in the study (monitoring, blood tests, ECGs, cardiac catheterization at baseline) can be considered as part of the optimal care of patient post-myocardial infarction. All, that is, but two. The investigator proposes to add a research intervention to a randomly selected 5% subset of patients on both arms of the study, namely, cardiac biopsy at the time of the first coronary angiogram and another cardiac catheterization with biopsy at six weeks post-MI. Although the added procedures entails some risk -- including a small chance of death as a result of the procedure, the investigator asserts that important information regarding the effect of the study agent on cardiac tissue may be gained by the procedure. As

well, she asserts, valuable information may be gained about the relationship between histological changes and the clinical course of patients post-MI. To reiterate, these procedures are without therapeutic justification; their purpose is to obtain more knowledge about cardiac disease in general.

In the design of the trial, mindful of the fact that men and women will likely respond differently to the proposed treatment, the investigator has restricted the study population to women. This will, assuming a fixed sample size, increase the validity of the study; many more subjects would be needed to prove the efficacy of the treatment in men as well in women. What has been gained in validity, though, has been at the cost of generalizability. The results of this study will only apply to a subset of persons with cardiovascular disease, namely, women. Freedman and Shapiro have observed that this trade-off is inherent in defining the eligible population of a clinical trial:

> One major choice is settled, knowingly or unknowingly, in setting eligibility criteria, that of validity vs generalizability. When a controlled trial imposes rigid criteria, the aim is to narrow the population to two groups similar in as many relevant respects as feasibly may be determined, differing only on the allocation of treatment of the two groups. Such a 'fastidious' trial is intended to present the cleanest possible scientific comparison. The choice though loses in generalizability what it may have gained in validity. The trial is intended to teach us

> something about the treatment of patients, who do not have the luxury of checking a list of eligibility criteria before choosing to become ill. The clinical goal of a trial -- can be sacrificed in 'fastidious' trials to an unyielding commitment to scientific validity.[11]

How does the restricted generalizability of the trial fit in to the ethical analysis of the study?

The ethical assessment of the generalizability of a study is a part of the assessment of its scientific value. Regarding the evaluation of the value of a study, Freedman has noted that

> [m]ost important, a judgement of value must include a view about the significance of the hypothesis itself: for reason of its novelty, clinical or other social implications, scientific interest, or otherwise.
>
> This would also be the occasion for broadly considering to what degree the proposal represents a duplication of established results or instead adds significantly to what the scientific community already knows.[12]

The generalizability of a study defines its "clinical or other social implications" and, thus, generalizability is a subset of scientific value. The results of a narrowly focussed clinical trial, aimed at only a subset of the patient population, will likely only have "clinical... implications" for that restricted population subset. Conversely, the results of a broad-based trial which includes a representative sampling of the patient population for a given medical disorder, may have substantial clinical

implications for the entire patient population. Thus, widely generalizable clinical trials have more scientific value that narrowly generalizable trials. In the Department of Health and Human Services Regulations, this assessment of the value of a study, and hence the assessment of its generalizability, is incorporated into the overall risk-benefit calculus. Namely, §46.111(a)(2) requires that IRBs ensure that

> [r]isks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result.[7]

Colleagues and I have argued elsewhere that the risks of therapeutic elements of clinical research must be analysed separately from the risks of interventions that carry no therapeutic warrant.[13] We have referred to the former as "therapeutic risk" and the latter as "dedicated research risk". Therapeutic interventions, as their associated risks, must satisfy the principle of clinical equipoise. As we point out,

> [f]or a nonvalidated intervention to be in equipoise with a standard treatment arm, its associated expectations of risk and benefit must be roughly equivalent to those of treatments used in clinical practice (or placebo if no treatment is commonly accepted).[13]

Thus, in this trial, the new therapy and the interventions done with therapeutic intent (blood tests, ECG, the initial

cardiac catheterization) must be in equipoise with standard therapy. This, as we have said, may be presumed to be the case.

Research interventions that present "dedicated research risk" must be weighed separately from therapeutic interventions. According to §46.111(a)(1), IRBs must ensure that dedicated research risks are minimized

> (I) by using procedures which are consistent with sound research design and which do not unnecessarily expose subjects to risk, and (ii) whenever appropriate, by using procedures already being performed on the subjects for diagnostic and treatment purposes.[7]

The IRB must furthermore ensure that these risks are reasonable in relation to "the knowledge that may reasonably be expected to result". Thus, the analysis of dedicated research risks does not involve a risk-benefit analysis, but rather a "risk-knowledge" analysis.

The "knowledge that may reasonably be expected to result" from a study is in part a function of the generalizability of a study. More broadly-generalizable studies produce results which are more important; narrowly defined, that is narrowly generalizable, studies produce results that are less important.

In our example, then, the IRB must ensure that the risks of the dedicated research interventions -- cardiac biopsy and cardiac catheterization -- are minimized and are

justifiable in relation to the importance of the knowledge likely to be gained. The dedicated research interventions in our example were intended to seem difficult to justify in the context of a study that is less than very important. (Let us at least assume this to be the case). Thus, the IRB may well decide that, in the context of a study that is not widely generalizable, such as one restricted to the study of one gender, these risks are unacceptable. Given this conclusion, two options are: (1) the interventions may either be eliminated from the study or, (2) if the investigator is unwilling to make this concession, the study as a whole may be disallowed.

A third option exists. Implicit in what we have said is that these dedicated research interventions may be allowable in a study which is either more generalizable or more important or both. Were the study to address the treatment of both men and women post-MI, the increased generalizability and importance of the study may be sufficient to allow the IRB to approve the dedicated research interventions. As we have said, though, this would require a much larger sample size. According to the discussion presented in chapter 3, to address the issue of gender in the primary analysis (with the same power, etc.), a total of 16000 subjects would be needed: 8000 men and 8000

women, with approximately equal numbers of each group assigned to the two treatment arms.

Note well though that in this case increasing the sample size of the study is more or less incidental to increasing the scientific value of the study. The ethical acceptability of the revised study derives from the fact that the study question has been changed: The initial study asked (with respect to the dedicated research intervention), "How do histological changes in the heart correlate with survival of women post-myocardial infarction?" The revised study asks a different question, namely, "How do these changes correlate with the survival of men and women post-myocardial infarction?" The latter study is acceptable not because it is a larger study, but rather because it asks a more important question.

With this example, I have attempted to demonstrate that considerations of the generalizability of a study can, and indeed ought, to be incorporated into the ethical analysis of research under current DHHS Regulations. Ethical considerations regarding the selection of subjects, of course, involve considerations regarding the equitable distribution of the burdens and benefits of a study. This example has attempted to highlight the fact that considerations regarding the generalizability of a study are

a subset of value-related considerations. As such, generalizability is weighed against dedicated research risks in the IRB's analysis of the risks of research. It follows that more dedicated research risk is allowable for studies that are widely generalizable -- that is, addressing a question that is of more scientific value -- than those that have narrowly defined study populations.

Although we have begun to characterize the relationship between generalizability and ethically acceptable risk in clinical research, questions remain as to the relationship between scientific value (broadly construed) and research risk. How ought the potential social impact -- apart from considerations of generalizability per se -- of a clinical trial be incorporated into its ethical analysis? Is the level of allowable dedicated research risk greater for trials involving diseases that affect more people in society than other diseases (say trials involving lung cancer versus those involving cancer of the cervix, or trials in HIV/AIDS versus gene therapy protocols for adenosine deaminase deficiency)? If this is the case, then persons who suffer from a common disease could be exposed to more risk without the prospect of therapeutic benefit in a clinical trial than those who suffer from rare diseases. Clearly, work is needed to resolve this issue.

Original contributions in this thesis

The Guidelines for Thesis Preparation allow persons submitting Masters thesis the option of indicating which, if any, portions of the thesis represent original contributions. The following portions of this thesis are original contributions:

- the historical overview of evolving ethical issues in the selection of subjects for clinical research (chapter one);
- the "meta-analytic" review of the literature presented in the second chapter which derived quantitative estimates of the proportion of patients excluded by protocol, physician and patient factors;
- the "meta-analytic" review of the literature presented in the third chapter which identified differences between trial participants and non-participants and which suggested that these differences were largely due to criteria for trial eligibility;
- the solution to the purported dilemma regarding proof of effect in subgroups versus the ethical requirement not to deny trial participants "proven" therapy when an effect has been demonstrated in the aggregate (discussion, chapter three);

and, finally,

- the analysis of the incorporation of generalizability into current understanding of the ethical analysis of research required by DHHS Regulations (chapter four).

References

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

2.Friedman MA, Cain DF. National Cancer Institute sponsored cooperative clinical trials. Cancer 1990; 65: 2376-2382.
3. Wittes RE, Friedman MA. Accrual to clinical trials. Journal of the National Cancer Institute 1988; 80: 884-885.

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

4. Svensson CK. Representation of American blacks in clinical trials of new drugs. JAMA 1989; 261: 263-265.

5. General Accounting Office. Women's health: FDA needs to ensure more study of gender differences in prescription drug testing. Washington, D.C.: Government Printing Office, 1992. (Publication no. GAO/HRD-93-17.)

6. Bird CE. Women's representation as subjects in clinical studies: a pilot study of research published in JAMA in 1990 and 1992. In: Mastroianni AC, Faden R, Federman D (eds.). Women and Health Research: Ethical and Legal Issues of Including Women in Clinical Studies (Volume 2). Washington, D.C.: National Academy Press, 1994: 151-173.

7.Department of Health and Human Services. Protection of Human Subjects. Title 45 Code of Federal Regulations Part 46: Revised June 18, 1991.

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

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

10.Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991; 266: 93-98.

11.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.

12.Freedman B. Scientific value and validity as ethical requirements for research: a proposed explication. IRB: A Review of Human Subjects Research 1987; November/December: 7-10.

13.Freedman B, Fuks A, Weijer C. Demarcating research and treatment: a systematic approach for the analysis of the ethics of clinical research. Clinical Research 1992; 40: 653-660.