Contemporary Clinical Trials 45 (2015) 261–264

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Some reflections on evaluating institutional review board effectiveness

David B. Resnik

National Institute of Environmental Health Sciences, National Institutes of Health, 111 Alexander Drive, Research Triangle Park, Box 12233, NC, 27709, United States

ARTICLE INFO

Article history:
Received 15 July 2015
Received in revised form 17 September 2015
Accepted 21 September 2015
Available online 25 September 2015

Keywords:
Institutional review boards
Effectiveness
Ethics
Oversight
Empirical research
Randomized controlled trials
Prospective cohort studies

ABSTRACT

Most commentators agree that it is important to conduct empirical research on the effectiveness of institutional review board (IRB) review and oversight but the studies that have been published so far do not directly address this question because they do not attempt to measure the impact of the IRB on the welfare or rights of human subjects. Additional studies on IRB composition, staffing, review times, consistency, and so forth will not yield that evidence that is needed to measure IRB effectiveness if they do not also collect data on the welfare and rights of human research subjects. Researchers should consider developing studies, such as randomized, controlled trials, or prospective, cohort designs, which directly measure IRB effectiveness. Such studies could yield information that will be invaluable in improving IRB review and oversight and protecting human research subjects.

Published by Elsevier Inc.

1. Introduction

Dozens of empirical studies have been published in recent years on institutional review board (IRB) review and oversight. Abbott and Grady have conducted a systematic review of this literature. [1] They distinguish between several types of studies: 1) studies of IRB structure, including descriptions of review times, costs, and IRB member characteristics; 2) studies of IRB process, such compliance with federal regulations; 3) studies of variation of IRB review in multisite research; and 4) studies of IRB decisions and deliberations. [1] Silberman and Kahn have published a useful review of the literature pertaining to IRB costs, review times, administrative burden, and variation in decision-making [2], and Nicholls and colleagues published a review of the theoretical frameworks and methodologies used to study IRBs. [3] All of these authors stress the importance of conducting additional empirical research on IRBs. In discussing a potential research agenda, Abbott and Grady ask some foundational questions pertaining to measuring IRB effectiveness:

> What do we expect from IRBs? Clarifying expectations is important to being able to measure effectiveness. For example, should we expect IRBs to be more consistent at determining the risk of certain procedures or the risk level of a study? In protecting subjects from risk, could we examine how IRBs minimize risk? Or how changes in study proposals required by the IRB protect participants from risk? Centralized data on the risks that research participants experience would also be helpful in this regard…Importantly, efforts should be made to identify and test metrics for measuring the quality of IRB review and the effect of IRB review on protection of human subjects. It may be necessary to first develop a consensus view on how IRB quality should be understood [1,pp. 7–8].

I agree with Abbott and Grady that we need to have a better understanding of what we mean by IRB effectiveness. To conduct useful and informative empirical research on IRBs, it is necessary to develop a conceptual framework that we can use to guide study design and interpret research results. We need to take a step back from our data collection on IRBs and ask some fundamental questions about what we are trying to measure and why.

2. The purpose of the IRB

To determine whether something is effective, we must have a prior understanding of its purpose, since effectiveness is a measure of how well something does what it is supposed to do. The first question to ask is "what is the purpose of an IRB?" The answer to this question may seem obvious to most readers but it sometimes gets lost in the discussions of costs, review times, and administrative burdens. To appreciate the answer to this question, a little bit of history is in order. An IRB-precursor emerged in 1965, when the National Institutes of Health (NIH) established a committee on its intramural campus to review selected research protocols for risks to human subjects prior to proceeding with experiments. [4] This review process became a formal requirement for all NIH-funded research involving human subjects in 1966. In 1971, the Food and Drug Administration (FDA) adopted a

E-mail address: resnikd@niehs.nih.gov.

http://dx.doi.org/10.1016/j.cct.2015.09.018
1551-7144/Published by Elsevier Inc.

similar requirement for research on drugs and medical devices submitted to the agency. [4]

In 1972, the Tuskegee Syphilis Study, in which 399 African-American men with untreated syphilis were enrolled in federally-funded research without their consent and denied treatment for syphilis when it became available in the 1940s, drew the attention of the national media and the U.S. Congress. In 1973, Congress held hearings on abuses of human subjects in federally-funded research and passed the National Research Act, which authorized federal agencies to develop regulations for research involving human subjects and created a national commission to study the ethics of conducting research on human subjects. [4] The National Commission for the Protection of Human Subjects in Biomedical and Behavioral Research released the Belmont Report in 1979, which described three ethical principles for conducting research on human subjects, i.e. respect for persons, beneficence, and justice. [4,5] The Belmont Report provided a conceptual foundation for a major revision of the Department of Health and Human Services (DHHS) regulations in 1981. The DHHS regulations, which are also known as the Common Rule because they have been adopted by 17 federal agencies, define IRB composition, authority, and function. [6] The FDA adopted similar regulations. [7] In 1991, DHHS added three sub-parts to the Common Rule to provide additional protections for pregnant women, fetuses, neonates, children, and prisoners. [6]

The Common Rule has remained largely unchanged since 1991, despite numerous calls for reform. [8,9] In 2011, the Office of Human Subjects Research Protections (OHRP), which oversees DHHS human subjects research, and the FDA, published an Advance Notice of Proposed Rulemaking (ANPRM), which includes proposals for reforming the research regulations to enhance protections for human subjects, reduce burden, delay, and ambiguity for investigators. [10] The agencies received thousands of comments from the public and federal agencies on the ANPRM. Sixteen federal agencies that have adopted the Common Rule agencies recently published a Notice of Proposed Rulemaking (NPRM) in response to comments received on the ANPRM. [11] The agencies need to respond to comments on the NRPM before revising the Common Rule. Although the federal regulations do not explicitly state the purpose of an IRB, it is clear from this brief historical review that IRBs were formed primarily to protect the welfare and rights of human research subjects. [12] These committees have taken on secondary purposes since then, such as facilitating research and protecting investigators, institutions, and sponsors from legal liability. [13]

3. Measuring IRB effectiveness

How might one determine whether IRBs are effective at protecting the welfare and rights of human research subjects? To answer these questions, we must define outcomes that can be equated with protection of welfare and protection of rights. This is not an easy task because "welfare" and "rights" are malleable concepts which are not easily quantified. For comparison, it is relatively easy to demonstrate whether the Environmental Protection Agency (EPA)'s air pollution regulations have been effective, because the regulations define acceptable ambient air concentrations for specific pollutants. [14] There are, however, no readily identifiable measures of welfare or rights which are analogous to protection of air quality. To measure IRB effectiveness, one must therefore develop some quantifiable measures pertaining to the welfare and rights of research subjects.

As Abbott and Grady point out, while research institutions keep track of adverse events and unanticipated problems, most do not routinely gather data pertaining to the welfare or rights of research subjects per se, so this information would need to be collected as part of a study. [1] Measurable indicators of welfare might include health-related outcomes pertaining to research participation, such as mortality, disability, morbidity, general health, and psychological well-being. One could collect data pertaining to these variables by asking research subjects to complete health surveys or questionnaires. One could also obtain information concerning the health of research subjects from their physicians, if they consent to the release of medical information, or from the investigators in charge of the studies in which they are enrolled. One could use the national death index to determine whether subjects have died during research participation. [15] Measures related to protection of human rights could include information about the informed consent process, confidentiality, privacy, and other issues related to respectful treatment, which could be obtained by interviewing or surveying research participants or investigators. To obtain accurate information, data collection might need to take place at different intervals during research participation and afterwards. Data collection might involve some complex logistical issues, since research participants and investigators probably would both need to consent to data collection and assist with it. Researchers could categorize data and aggregate it into an overall measure of welfare or respectful treatment for each subject.

4. Measuring variables that may impact IRB effectiveness

In addition to collecting data concerning primary outcomes (i.e. welfare and respectful treatment), researchers could also collect data on a variety of factors related to research review and oversight that could potentially impact these outcomes, such as:

- Number of protocols overseen by the IRB
- The types of studies reviewed by the IRB, e.g. clinical trials, epidemiological research, social/behavioral research, etc.
- IRB composition, e.g. percentage laypersons, scientists, and independent members on the IRB; experience and expertise of IRB members
- IRB review times
- Total number of IRB staff and staff per protocol
- Total costs of IRB review and costs per protocol
- Frequency and quality of auditing activities conducted by the IRB or designees (e.g. quality assurance officers)
- Frequency and quality of educational activities on human research protections conducted by the IRB or the institution
- Accreditation status of the IRB
- IRB compliance with regulations and guidelines
- IRB decision-making process
- Internal and external consistency of IRB decisions
- Use of centralized IRB review for cooperative research activities
- Organizational justice, i.e. perceived fairness of relationships between the IRB and investigators.

It is worth noting that the studies included in the literature reviews of research on IRBs do not directly measure IRB effectiveness because they do not collect data on subject welfare or rights. [1–3] These studies only collect data on potential predictors of IRB effectiveness (or surrogate endpoints), such as the factors listed above. Surrogate endpoints related to an intervention may or may not reliably predict its effectiveness. For example, a new cancer drug might reduce tumor size without increasing longevity or improving quality of life. [16] Although it is important to gather data on variables that may predict IRB effectiveness, such studies are no substitute for research that measures actual effectiveness.

It is also worth noting that IRB costs and some other variables may be inversely related to effectiveness. An IRB might have quick review times and minimal administrative burdens and other costs, yet be ineffective at protecting human subjects. For example, Coast IRB, a private company located in Colorado, advertised that it offered quick review times at minimal costs. The IRB was caught in a Congressional sting operation for approving a fake protocol for studying a fictitious adhesive gel for use in abdominal surgery. The protocol did not clearly identify the ingredients of the gel, which would constitute a significant safety issue. Two IRBs rejected the protocol as unsafe but Coast approved it