Journal of Clinical Research Best Practices

Vol. 17, No. 7, July 2021 "Happy Trials to You"

Standards of Evidence for IRB Decision-Making

By David B. Resnik

Introduction

Institutional review boards (IRBs) review clinical studies and approve those that meet regulatory criteria and ethical guidelines. The review process considers various factors, such as whether the generalizable knowledge likely to be generated by the study justifies the risks of harm imposed on study participants. Because clinical studies are experiments, there are many unknowns. In all but the most trivial cases, IRBs must, therefore, deal with uncertainty.

In a criminal case, there must be sufficient evidence to support the decision to convict. Without enough evidence, the court must find the defendant "not guilty," which is not the same as "innocent." Similarly, when an IRB disapproves a study, it is not necessarily saying it is a bad study; it is just saying it does not have enough evidence to say the study is worthy of approval.

When an IRB says it has enough evidence to support approval, it is not necessarily saying that it is 100 percent confident in its decision. It is just saying that it is confident enough, where the word "enough" implies a minimum and often unstated level of confidence in the decision (e.g., 75 percent). This level of confidence is the key element in a standard of evidence.

Standards of Evidence

A standard of evidence establishes a level of certainty that something has been demonstrated to be true. In a clinical study, the typical standard of evidence is a probability of less than five percent that an observed difference could have occurred just by random chance (i.e., $P \leq 5\%$). Other statistical measures, such as paired T-tests, Chi-square tests and linear regression can also be used.[1]

U.S. federal regulations (i.e., the Common Rule, 45 CFR 46, and the Food and Drug Administration (FDA) Regulations, 21 CFR 56) define the IRB's role in the oversight of research with human subjects, addressing such matters as the IRB's composition, functions, procedures, responsibilities and review criteria. The regulations do not, however, define a standard of evidence for IRB decision-making and leave considerable discretion to individual IRBs to decide whether a study should be approved, disapproved, revised, temporarily suspended or halted.[7] Instead, they implicitly leave it to each IRB to establish its own standard of evidence (or not).

The absence of a standard of evidence for IRB decision-making has important ramifications for the oversight of clinical research. For one thing, it can lead to inconsistency in IRB review. One IRB might, for example, decide that the evidence is strong enough to approve the study, while another IRB might decide that the evidence is too weak.[7] Similarly, when presented with two equivalent studies, an IRB may approve one and disapprove the other.

Inconsistent decision-making is problematic for two reasons. First, it can lead to the perception that IRB decisions are arbitrary and therefore unfair. The perception of

unfairness can demoralize investigators and damage organizations. It can also increase the risk of noncompliance with regulations and policies.[5]

Second, inconsistency can be an indicator of substantive problems with IRB review. The IRB may be under-protecting or overprotecting study participants. It may even be doing both in different studies. Under-protection occurs when an IRB approves a study based on inadequate evidence, whereas overprotection occurs when an IRB disapproves a study despite adequate evidence that it meets the approval criteria. While under-protection can threaten the rights and well-being of study participants, overprotection can impede valuable scientific research.[7]

A substantial body of research indicates that different IRBs often make conflicting decisions concerning the same study.[3,4,6] For example, in 2002, Lee Green and collaborators submitted a health services research protocol to 43 IRBs at Veterans Affairs clinics around the U.S. Ten IRBs approved the study through an expedited review process; 31 approved the study through full board review, one decided the study was exempt from IRB review; and one decided the study was too risky to approve.[3]

A plausible explanation for the inconsistency in IRB decision-making is that IRBs interpret and apply the regulations differently, due to differences in knowledge and experience, education and training, and cultural and moral values.[2,7,9] This variability may also be due, in part, to differences in standards of evidence used by IRBs.

Should all the members of a given IRB use the same standard or should each member be allowed to use his or her own standard? Should the IRB apply the same standard of evidence to all the studies it reviews? In the absence of regulations or guidance from regulatory agencies, what standard of evidence should IRBs use? These are important questions to ponder, but, surprisingly, very little has been written about them until now.[7,10]

Should all the members of a given IRB use the same standard of evidence or should each member be allowed to use his or her own standard?

A strong case can be made that all members of an IRB should use the same standard of evidence. Allowing IRB members to use different standards could undermine deliberation and decision-making by leading to perplexing disagreements that the IRB members do not even understand themselves. IRB members may agree on all the facts and expert assessments, but still disagree on the conclusion because they are, perhaps unknowingly, using different standards in evaluating the facts and assessments.

Should IRBs apply the same standard of evidence to all the studies it reviews?

IRBs should employ a standard of evidence appropriate to the study under review. For example, a study that imposes great burdens and risks on the participants may call for the beyond reasonable doubt standard. At the other extreme, expedited review with the preponderance of evidence standard is probably adequate for a study that imposes minimal burdens and risks.

In the absence of regulations or guidance from regulatory agencies, what standard of evidence should IRBs use?

An IRB could use many different types of standards of evidence, including standards from science, technology, engineering and mathematics (STEM), and law. While standards of evidence used in STEM fields are rigorous and reliable ways of establishing empirical knowledge, there are at least two reasons why IRBs should use legal standards. The first reason is that IRBs include lay members who are not scientists, engineers, mathematicians

or health professionals. It takes years of education, training and practice to learn how to use standards of evidence from STEM fields. Lay members usually do not have this background, and such expertise may not be the strong suit of other members. If the IRB should use one standard of evidence, it should use a standard that can be understood by all members, not just those members with the required education and expertise. Legal standards satisfy this requirement because they have been developed for jurors in criminal and civil trials.[7] While they are not nearly as rigorous or reliable as the rules and procedures that apply to STEM fields, they are time-tested and easily understood.

Another reason for using legal standards is that the standards of evidence used in STEM fields apply to determinations that are descriptive or factual in nature, such as whether an observed difference between two treatments is statistically significant. Standards of evidence used in STEM fields can produce evidence (e.g., the likelihood of a particular adverse effect) for IRB consideration, but IRBs cannot use them to decide whether risks are justified or reasonable in relation to benefits because deciding whether something is reasonable is a qualitative, value-based judgment.[7] In contrast, legal standards are specifically designed to help jurors assess evidence and make qualitative, value-based judgments, such as whether someone's conduct was negligent or whether the terms of a contract are unfair.[7] Juries and IRBs are responsible for assessing evidence, not producing it.

Legal Standards of Evidence

Assuming that IRBs should use a legal standard of evidence, which standard should they use? The three most commonly used legal standards are: preponderance of evidence, beyond reasonable doubt, and clear and convincing evidence.[7] These standards involve making qualitative judgments concerning the strength of the evidence.

The least demanding standard is "preponderance of evidence" (POE), which means that the conclusion, given the evidence, is more likely true than false (i.e., probability > 50%). POE is used in civil lawsuits related to such matters as torts (e.g., negligence), contracts, patents and property. A probability of 51% percent means the decision will be correct a bit more than half the time.

The most demanding standard is "beyond reasonable doubt" (BRD), which means that the conclusion, given the evidence, is so likely to be true that no reasonable person would doubt it (i.e., probability ≥95%).[11] BRD is used in criminal cases because so much is at stake for the accused. A probability of 95 percent means that the decision will be correct 95 out of 100 times.

In between these two standards is the "clear and convincing evidence" (CEE) standard, which is used in cases where important human rights are at stake, such as matters involving child custody or involuntary commitment. Evidence is clear and convincing if it demonstrates that the conclusion is substantially more likely true than false (i.e., probability > 75%).[7,8] A probability of 75 percent means the decision will be correct three out of four times.

Which legal standard of evidence should an IRB use?

Although a study may call for the BRD standard (probability > 95%), the available evidence is unlikely to be adequate to meet that standard. Even well-designed clinical studies may fail to produce useful knowledge for various reasons, such as difficulties with recruitment or implementation, or failure to translate the results into medical practice.[7]

The POE standard (probability > 50%) sets too low a bar for most phase 1 to 4 clinical studies, since significant human rights can be at stake.[7] If approval will be the correct decision only 51 percent of the time, why not settle for 50 percent and just flip a coin?

As noted earlier, the CCE standard (probability > 75%) is used in legal matters other than criminal cases where significant human rights are at stake. One could argue that the CCE standard should also apply to IRB decisions for this reason.[7] One could also argue that a probability of 75 percent is too low to adequately protect study participants or that it sets an unrealistically high bar for the IRB and may therefore impede valuable research.[10] Nevertheless, given all the uncertainties IRBs deal with, it seems like a reasonable compromise.

IRBs do not review each study in a vacuum. Rather, they review a series of studies from which they can learn. If an IRB keeps track of its correct and incorrect decisions, it can use Bayesian probability to adjust its standard of evidence based on its track record. For example, if safety issues proliferate in approved studies, it can increase its standard of evidence. On the other hand, if it regularly disapproves studies that other research sites successfully conduct without incident, it can lower its standard of evidence.

Whatever standard a research site adopts, it should keep in mind that most IRB members do not think in terms of quantitative measurements of evidence, such as probabilities, especially when comparing, for example, 70 percent vs. 75 percent. Humans are wired to think in qualitative terms like "clear and convincing evidence." Two IRB members might debate whether the evidence is clear and convincing, but not whether it is more than 75 percent likely to yield a correct approval decision.

Conclusion

Lack of regulation concerning standards of evidence has contributed to inconsistent decision-making among IRBs, which is a symptom of deeper problems. Oversight agencies should consider providing guidance. In the absence of such guidance, IRBs should make their standard(s) of evidence explicit so members can better understand their own thinking, deliberate within a common understanding and communicate clearly to investigators why they need stronger evidence for approval when the IRB demands it.

The topic of evidentiary standards for IRBs is a complex and emerging issue that merits research: What evidence do investigators provide IRBs? How do IRB members evaluate evidence? What standards of evidence are implicit in their approvals? How confident are they in their judgments and conclusions? How accurate are their approvals and disapprovals?[7]

Acknowledgments

This research was supported by the Intramural Program of the National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH). It does not represent the views of the NIEHS, NIH or the U.S. government. This article is based on a previous article by the author.[7]

References

1. Weiss, P.A., *Introductory Statistics, 9th Ed*. (Pearson, Upper Saddles River, N.J., 2011).
2. Friesen, P., Yusof, A.N.M., Sheehan. M., "Should the Decisions of Institutional Review Boards Be Consistent?" *Ethics & Human Research,* 41(4):2–14. (2019)

3. Green, L.A., Lowery, J.C., Kowalski, C.P., Wyszewianski, L., "Impact of Institutional Review Board Practice Variation on Observational Health Services Research". *Health Services Research,* 41(1):214–230. (2006)
4. Hirshon, J.M., Krugman, S.D., Witting, M.D., Furuno, J.P., Limcangco, M.R., Perisse, A.R., Rasch, E.K., "Variability in Institutional Review Board Assessment of Minimal-Risk Research," *Academic Emergency Medicine,* 9(12):1417–1420. (2002)
5. Martinson, B.C., Crain, A.L., De Vries, R., Anderson, M.S., "The Importance of Organizational Justice in Ensuring Research Integrity," *Journal of Empirical Research on Human Research Ethics,* 5(3):67-83. (2010)
6. Petersen, L.A., Simpson, K., Sorelle, R., Urech, T., Chitwood, S.S., "How Variability in the Institutional Review Board Review Process Affects Minimal-risk Multisite Health Services Research," *Annals of Internal Medicine* 156:728–735.
7. Resnik, D.B., "Standards of Evidence for Institutional Review Board Decision-making," *Accountability in Research,* Dec 8, 2020:1-28 [online ahead of print].
8. Rothstein, P., Rader, M., Crump, D., *Evidence, 6th Ed* (West Publishing, St. Paul, Minn. 2010).
9. Silberman, G., Kahn, K.L., "Burdens on Research Imposed by Institutional Review Boards: The State of the Evidence and Its Implications for Regulatory Reform," *Milbank Quarterly,* 89(4):599-627. (2011)
10. Smith, E., Anderson, E.E., "Reimagining IRB Review to Incorporate a Clear and Convincing Standard of Evidence. *Accountability in Research,* February 8, 2021: 1-8 [online ahead of print].
11. Weinstein, J.B., Dewsbury, I., "Comment on the Meaning of Proof Beyond a Reasonable Doubt.'" *Law, Probability and Risk,* 5:167–173. (2006)

Author

David B. Resnik, JD, PhD, is a bioethicist and IRB chair at the National Institute of Environmental Health Science, National Institutes of Health. Contact him at resnikd@niehs.nih.gov.