



Published in final edited form as:

AJOB Prim Res. 2011 ; 2(2): 24–33. doi:10.1080/21507716.2011.601284.

The Myth of Community Differences as the Cause of Variations Among IRBs

Robert Klitzman

Columbia University

Abstract

Background—Although variations among institutional review boards (IRBs) have been documented for 30 years, they continue, raising crucial questions as to why they persist as well as how IRBs view and respond to these variations.

Methods—In-depth, 2-hour interviews were conducted with 46 IRB chairs, administrators, and members. The leadership of 60 U.S. IRBs were contacted (every fourth one in the list of the top 240 institutions by NIH funding). IRB leaders from 34 of these institutions were interviewed (response rate = 55%).

Results—The interviewees suggest that differences often persist because IRBs think these are legitimate, and regulations permit variations due to differing “community values.” Yet, these variations frequently appear to stem more from differences in institutional and subjective personality factors, and from “more eyes” examining protocols, trying to foresee all potential future logistical problems, than from the values of the communities from which research participants are drawn. However, IRBs generally appear to defend these variations as reflecting underlying differences in community norms.

Conclusions—These data pose critical questions for policy and practice. Attitudinal changes and education among IRBs, principal investigators (PIs), policymakers, and others and research concerning these issues are needed.

Keywords

attitudes; culture; ethics; qualitative research; policy; regulation; research ethics

Variations among institutional review boards (IRBs) have been documented for almost 30 years (Goldman and Katz 1982) and can impede research, yet continue (McWilliams et al. 2003; Stair et al. 2001), posing critical questions as to why as well as how these differences should best be addressed.

Since the regulations governing IRBs (U.S. Department of Health & Human Services 2009) were developed over three decades ago, the amount of multicenter, as opposed to single-site, research has increased dramatically, generally requiring multiple IRBs to review a protocol.

A review of 40 peer-reviewed articles on IRB variations (Greene and Geiger 2006) found that IRBs often vary in length of time to approval (Vick et al. 2005); type of review (i.e., exempt, expedited, or full review) (Dyrbye et al. 2007); and type and number of changes in consent forms (Stark, Tyson, and Hibberd 2010), which can affect rates of refusal by potential participants (Dziak 2005) and the length and readability of consent forms. Multiple reviews delay the initiation of studies, and require resources that could otherwise be spent on the scientific research itself (Green et al. 2006). These variations can make comparison of data collected at different sites difficult, if not impossible. Important research can thus be delayed or hampered (Burman et al. 2001; Dyrbye et al. 2007; Dziak et al. 2005; Green et al. 2006; Jamrozik and Kolybaba 1999; Middle et al. 1995; Newgard et al. 2005).

The U.S. regulations (45 CFR 46) recognize the potential roles of relevant local and state laws and sensitivity to “community attitudes”:

The IRB shall be sufficiently qualified through the experience and expertise of its members, and the diversity of the members, including consideration of race, gender, and cultural backgrounds and sensitivity to such issues as community attitudes, to promote respect for its advice and counsel in safeguarding the rights and welfare of human subjects. ... If an IRB regularly reviews research that involves a vulnerable category of subjects, such as children, prisoners, pregnant women, or handicapped or mentally disabled persons, consideration shall be given to the inclusion of one or more individuals who are knowledgeable about and experienced in working with these subjects. (U.S. Department of Health & Human Services 2009)

IRBs must also include a nonaffiliated member (often anecdotally called the “community” member).

But surprisingly, no data have ever been published supporting the claim that inter-IRB differences in fact reflect varying community norms, and studies documenting variations (McWilliams et al. 2003; Stair et al. 2001) have not examined *why* exactly IRBs differ in their decisions (Greene and Geiger 2006). Questions concerning whether these inconsistencies reflect variations in community values and/or other factors, and if so, when and how, have not been probed.

In a qualitative, in-depth, semistructured interview study I recently conducted of views and approaches toward integrity of research and related issues among IRB chairs, directors, administrators, and members (Klitzman 2011a), issues concerning variations among IRBs repeatedly arose. The study aimed to understand how IRBs viewed research integrity (RI)—which these participants defined very broadly—and related areas. Interviewees discussed how they defined, viewed, and addressed RI, based in part on how they interpreted and applied federal regulations, which they did in differing ways due to a variety of factors. Interviewees varied in how they saw and approached the roles and responsibilities of IRBs and conflicts of interest (Klitzman 2011b); how they viewed and interacted with researchers, federal agencies, institutions, industry funders, central IRBs (Klitzman 2011c), and developing-world IRBs (Klitzman 2011d); and how they were affected by histories of violations of RI and audits at their own and other institutions, and psychological and

personality issues on their IRB. They described how and why their IRBs differed in these and other regards. The qualitative methods used in the study allowed for further detailed explorations of domains that emerged and shed additional light on these issues. Given the growing concerns about such issues, and particularly about IRB variations, this paper explores the nature, scope, causes of intra-IRB differences, and possible alternative ways of addressing these.

METHODS

As described elsewhere, I interviewed 46 chairs, administrators, and members by telephone for 2 hours each (Klitzman 2011a; Klitzman 2011b; Klitzman 2011c; Klitzman 2011d). In brief, from the list of the top 240 U.S. institutions by National Institutes of Health (NIH) funding, I e-mailed the IRB leaders of every fourth one ($n = 60$). Of these institutions, I recruited IRB leaders from 34, yielding a response rate of 55%. At times, I interviewed an institution's chair/director and administrator (for instance, if the chair suggested that particular questions might best be answered by the latter). Thus, I recruited 39 individuals from these 34 institutions, which varied in geography, size, and public/private status. In addition, I asked half of the IRB leaders (every other one) to disseminate a flyer about the project to IRB members to recruit one member, too, from each IRB. I conducted seven additional interviews (with one community member and six regular members).

The interviews explored perspectives on integrity of research, defined broadly, responses of IRBs to problems in these areas, and factors entailed, and also illuminated other, related themes that emerged regarding IRBs' decisions and interactions with researchers and each other, and centralized IRBs. The appendix presents sections of the semistructured interview guide, designed to gain detailed descriptions of these issues. When examining people's lives and decisions, Geertz (1973) argued that it is crucial to obtain a "thick description" to comprehend these individuals' experiences from their own perspectives, rather than by imposing external theoretical structures.

The methods employ elements from grounded theory (Strauss 1990), as described elsewhere (Klitzman 2011b).

In data analyses, I triangulated methods, informed by the published literature outlined earlier. Interview transcription and initial analyses occurred during the time frame when I was conducting the interviews. The study was approved by the Columbia University Department of Psychiatry Institutional Review Board, and subjects all provided informed consent.

After completion of all interviews, I conducted further data analyses in two phases, with a trained research assistant (RA). In phase I, we each independently analyzed a subset of transcriptions to examine factors that affected interviewees' IRB experiences, noting categories of recurrent themes to which we gave codes. We systematically read each transcription, coding passages to assign "core" codes or categories (e.g., perceptions of differences among IRBs). We each inserted a topic label (or code) next to each excerpt in the transcription to describe the themes discussed. Together, we subsequently reconciled

these coding schemes, developed independently, into a single scheme. Then we developed a coding manual, defining each code, and exploring disagreements until reaching consensus. When we identified new themes that did not fit into the initial coding manual, we discussed these and modified the manual as appropriate.

We then, in the second phase of the analysis, independently content-analyzed the data to identify the main subcategories, and variations within each of the primary codes. We reconciled the subthemes we each identified into a single group of “secondary” codes, and an elaborated group of primary codes. These subcodes included specific types of differences in interpretations (e.g., reasons why IRBs approached issues differently).

We then used these codes and subcodes, and both analyzed all the transcriptions. We probed disagreements through further analysis until arriving at consensus, and regularly checked for consistency and accuracy by comparing earlier and later coded passages of text. We systematically developed and documented the coding processes.

RESULTS

Table 1 (Klitzman 2011c) summarizes the sociodemographics of the sample.

Overall, as outlined in Figure 1, and described more fully later in this paper, interviewees often discussed variations among IRBs, reporting differences in how IRBs viewed and approached critical aspects of study protocols. Interviewees generally argued that these differences reflected community values and thus were justified. But when probed in the interviews, these variations tended to appear instead to reflect other personality or institutional factors. IRBs appear to be engaged in complex, subjective human processes. Particular aspects of institutional histories (e.g., past federal audits and “shutdowns” of research) and IRB chairs’ and/or vocal members’ personalities, idiosyncrasies, and abilities to anticipate all future logistical problems in carrying out multifaceted protocols, more than differences in community values, frequently seemed to underlie these inconsistencies.

Types of Variations

IRBs often disagreed in interpreting and defining key terms in the federal regulations. The regulations do not fully clarify these terms, leaving IRBs to do so on their own, in varying ways. Yet these terms embody inherent ambiguities. In part because they must place all research into one of several categories (e.g., expedited review), IRBs often face difficulties in defining “minimal risk,” “no more than minor increase over minimal risk,” and “research.” As presented later, perceptions of “risk” ranged from direct to indirect, and from likely to unlikely. Interviewees described how their subjective concerns often affected assessments of the likelihood and seriousness of harms.

Frequently, IRBs also felt that, on ethical grounds, they had a role in the scientific design of protocols, if the science was not “good enough” (i.e., the possible social benefit is insufficient to justify the risks). But definitions of this standard (i.e., “how good the science needs to be”) can vary as well. Interviewees reported wrestling with questions, too, of at what point to stop making small changes in consent forms—when “enough is enough.”

In making these assessments, IRBs see themselves as varying in “flavors” and “colors,” and in how “nitpicky” versus “user-friendly” or “flexible” they are. Differences among IRBs appear to not simply follow a clear or consistent dichotomy of “liberal” versus “conservative.” As one chair said,

Each of our IRBs has its own feel or flavor. That’s good. The issues tend to be how picky one IRB is going to be vs. the other. Some will table a protocol for a typo. Others will give specific minor revisions. It’s rare that a protocol gets tabled by one IRB and fully approved by another, but it is interesting when issues keep coming up at one IRB, but not another. (IRB3)

Locations of Variations

Interviewees described variations as occurring even within a single institution and in a single IRB over time.

Disagreements within one institution—Many institutions have more than one IRB, which can disagree about a protocol that is mutually reviewed. Thus, even within one institution, IRBs may be inconsistent.

Different IRBs at our institution feel totally different. One chair is very obsessive and particular. His IRB can spend an hour going over a consent form. In the IRB *I* run, everyone participates, and says what they want. We use a lot of the lay people to really help with “is this a useful consent form?” We say, “This sounds fine.” Then, *they* say, “But I don’t know what this word means.” So, we very much use our lay people, and empower them a lot more than some of the other IRBs do. It’s whether you’re going to be a committee that *really* discusses things, or a “Robert’s Rules of Order” kind of committee. We have both, and sometimes when we don’t have a quorum, people move around, and are on other IRBs, and are astounded that we’re at the same institution. (IRB3)

This chair suggests differences at his institution in both the process and content of IRB decisions, and distinctions between the letter and the spirit of the regulations. As discussed later, respondents tend to defend such variations, generally seeing their own IRB’s approach as justified, and implicitly somehow better (e.g., “we really discuss” issues and “empower” unaffiliated members, while another chair is “obsessive”). Yet at times, the same protocol may involve two principal investigators (PIs) in the same academic medical center, who submit it to two different IRBs, which then differ in their decisions.

Within some institutions, chairs occasionally meet to try to reduce these intra-institutional variations, but such efforts do not always occur or fully succeed.

A steering committee of the IRB chairs meets monthly, so there is some oversight, and attempt to get some consistency, but not much. (IRB3)

When chairs communicate or get together, differences can persist—for example, one chair may remain concerned about particular risks, and have much sway. As one vice-chair said:

When the chairs get together, there are definitely differences. The primary chair of my committee is very concerned about pain and suffering—*very* concerned—and

really takes it to heart when there's the slightest indication that subjects may experience what he determines as suffering. It's caused a lot of problems. (IRB11)

IRBs may thus have strong concerns about particular areas due to differences in the chairperson, not the community per se. But interviewees tend to defend such inconsistencies. Another chair explained that at times at her institution,

... three different PIs submit the same protocol. ... The protocols may not all go to the same committee, and we're really not consistent. The investigators get very upset: "How come you let so-and-so do it here, and they didn't have to change the consent or do this?" We've decided to at least try to be consistent. But we don't want them to be totally consistent. (IRB17)

She continues, justifying these differences.

Here in a big city, we are much more comfortable than in more conservative regions with certain issues that aren't a big deal for us to talk about, but I'm sure would raise eyebrows in other places, and be more difficult. So it's to be expected, and not a bad thing to have some variation based on community standards. (IRB17)

However, she then points out that divergent opinions emerge over time even from the same IRB.

But even our own committee isn't consistent all the time, depending on who is and isn't there that day, and how discussions go. Sometimes, somebody will get hung up on a small thing, and it will be blown out of proportion. Get enough people going on it, and it will become more of an issue than it should. (IRB17)

In a few instances, certain differences may indeed exist among communities. For instance, large cities may differ from isolated rural regions in attitudes toward HIV studies among gay men. But it is not clear that community attitudes among these large cities differ significantly. Moreover, in most research, even such variation (e.g., concerning attitudes related to HIV research among gay men) does not seem to occur.

Importantly, however, the preceding interviewee's justification (being in "a big city") does not fit the case she describes of differences within the same community. Specifically, all of the IRBs at her institution are located in the same "big city," yet at times disagree in substantial ways. Hence, the interinstitutional differences she describes appear to be related not to the local community (being in an urban vs. other area), but to the personalities and preferences of individual members. Still, she readily accepts, and does not seem very bothered by, these differences.

Such variation may also occur when a recently appointed chair or member raises new issues about a protocol that had previously been approved. The protocol may suddenly be challenged and not be reapproved. Another interviewee said,

Along comes a new associate chair like me. I'll question something, and the PI sends me back a nasty e-mail: "Why is this not being approved now? It was always approved in the past!" I think that's good. Because you always want fresh ideas. For example, a hematologist has recently joined our committee. None of us had

thought of bleeding with a specific drug, because none of us knew enough about thrombosis. We didn't really think about it. She did. So, if this protocol had come through before, it might have gone through as smooth as silk. But now, with this person on the committee, it was a stumbling block. That's OK. (IRB27)

Thus, IRBs may not only not be bothered by such differences, but may in fact see these as beneficial. Yet in this instance, the change results from additional scientific expertise on an IRB, which presumably should have been sufficiently present heretofore, and does not reflect differences in community values.

Interviewees may recognize that these variations suggest subjectivity in the review process, but argue that protocols have to be reviewed individually on a case-by-case basis, further legitimizing these differences. Nonetheless, IRBs tend to see their decisions, once made, as definitive. The interviewee just quoted continued,

It sounds arbitrary, but has to be taken case by case. ... Practicing medicine is the same thing. We try to be consistent, and have therefore developed lots of guidelines, but every protocol is so different, that flexibility is built in. (IRB27)

But, as we have seen, inconsistencies arise regarding not only different protocols, but the *same* protocol as well, suggesting that the cause of differences is not always the nature of the "case" or protocol—but rather the particular individual reviewers.

Inconsistency in an individual member—Not only a single IRB, but a single member may also make seemingly inconsistent decisions on similar protocols over time. A person may judge each protocol on its own, but then express seemingly contrasting concerns about different studies.

In our committee, one member is one of the fiercest critics of IRBs running amok, and constantly beating down investigators, and coming up with new things. He says we focus on trivial minutiae, and are blocking progress. But frequently he will also call for a test of comprehension, or doing extra things. ... He comes to a conclusion based on the case at hand. ... Even within this one guy, there is a pretty wide range. (IRB40)

Questions emerge here though, too, of how to define "inconsistencies"—that is, whether they stem from differences in definitions versus applications of IRB procedural versus content issues. Inconsistencies may entail differing views of the appropriate extent of an IRB's input or definitions of more abstract, philosophical concepts such as "justice" or "autonomy." Nonetheless, the overarching point here is that protocols can be reviewed differently due to a range of factors.

Causes of Variations

As mentioned earlier, inconsistencies frequently appear to arise due to differences not in community values, but in institutional and individual psychological factors, and abilities to foresee all possible logistical problems. A few interviewees recognize that differences in community values only partly accounted for variations.

The *stock answer* is to say that the IRB's supposed to consider community conditions, and that it's possible that what we do here and what you do there would be different—because of the population, because of who we are as researchers, the community setting, etc. That's why we have community members. To some extent that's a *piece* of it. (IRB22)

But other factors clearly emerge as well.

Types of institutions—At times, IRBs' decisions are shaped by histories and fears of federal “shutdowns of research” at institutions, which can contribute to different levels of anxiety about certain issues. Institutional size may also play a role. A small institution's IRB may be more flexible because IRB staff personally know the PIs, and can rely on ongoing trust and relationships that larger institutions may lack. Thus, a small institution can potentially be less rigid. Conversely, IRBs at larger, research-intensive medical centers may have developed more formalized policies and procedures, and have adhered to these more frequently (given the larger overall volume of research).

Because we're small, we can be fairly nimble. Our investigators collaborate with large university hospitals whose IRBs are pretty rigid about the format of their consent forms, which sound like they're written for very serious drug trials, when, in fact, a researcher is just doing a survey about tobacco smoking. At one hospital, if you change the format—just the way the consent form looks on paper—you have to get it reviewed. Because we're small, I just say, “Well, you changed the way it looked, and added the little logo there. Just give me a copy for my files.” (IRB26)

Again, causes of variation may thus emerge from differences in institutional cultures that do not necessarily reflect differences in laws, regulations, or attitudes of the communities from which study participants are recruited.

An IRB's past experiences—An IRB's past experiences can also shape how it views and approaches a protocol. IRBs differ in the degrees to which they have reviewed certain types of research. Frequently, an IRB may have difficulty initially, when assessing a new type of study, but then become more comfortable with the design or topic over time. An IRB's experience level can thus either facilitate or impede a review.

IRBs may implicitly weigh the riskiness of a particular study in relation to other research they have approved. An IRB's comfort may thus be shaped in part by the highest risk the IRB has previously accepted:

One IRB at a major cancer center doesn't view research on smoking prevention, psychological interventions, or prevalence. They don't see this research as worth worrying about, because from their point of view, the level of risk is so small compared to the rest of what they're doing. It hardly gets a yawn. (IRB22)

Conversely, certain IRBs or members may be less comfortable with social science research because they review it less.

I've seen hospital IRBs overreact, and get incredibly conservative and careful when reviewing a behavioral study, because they just don't really know how to treat it. It's new and different. (IRB26)

Psychological factors—Chairs' and/or vocal members' subjective beliefs, anxieties, and comfort may affect decisions as well. Many chairs and members have been assigned to the IRB without having had any formal ethics training, and rather than engaging in careful "ethical analysis," per se, draw on their "gut feelings" (IRB15) and "intuition," seeking "peace of mind" about the riskiness of a study. IRB members vary in how much they weigh their anxiety. Interviewees occasionally use "the sniff test" to sense whether a protocol seems ethically sound.

On an IRB, particular individuals may be overly cautious, "obsessive," and more or less flexible or strict. Members may have different standards, and some "tighter" people may argue that studies should be perfect, or not proceed.

A couple of members tend to come down quite hard on clinical research operations, and shut down more studies if issues don't look really fixed: "If we can't guarantee that studies are absolutely perfect, we should stop them." I think studies are never going to be absolutely perfect. There's no such thing as an error free human system ... one MD and one community person say, "Why don't you halt studies until they are completely fixed?" They are strident about saying you've got to cut down on all of this research until things are substantially or completely fixed. (IRB14)

These tensions highlight underlying questions of how to define "good enough," and how IRBs do or should do so.

At times, chairs and members rely, too, on their personal impressions of PIs, and vary in cautiousness, obsessiveness, or strictness. Members may each have their own "pet peeves." A member who is "the only oncologist"

may be the only person who has little nagging irritations about the subtleties of what constitutes a weak comparator of a chemotherapy regimen. Lots of many other nagging irritations come up from other people's perspectives. The community members will often come up with something that the medical people just don't think of. (IRB14)

As another example of one member having a particular concern, "One member will call the phone numbers on the consent forms to make sure they work. Sometimes they don't!" (IRB27)

Such concerns, though expressed by only one or a few members, can nonetheless affect the group's decision. Chairs, in particular, can play decisive roles in shaping each IRB. Differences may result due to an IRB's composition, and group processes and dynamics. As another interviewee elaborated,

IRBs in our institution vary, depending on the make-up of each individual committee. Different individuals are on each one. Some IRBs here focus a great deal on detail. Others are much more "big picture." (IRB11)

Group consensus can often, but not always, defuse a member's nitpickiness, but at other times, one obsessive chair or vocal member can still hold sway. Particular individuals, and hence IRBs, are seen as being more or less tough and "hard-nosed" or flexible in reviewing studies.

"Good catches": The effects of "many eyes" reading a protocol—IRBs can also diverge due to complicated logistics of studies, the details and implications of which may not be readily foreseeable on initial reading. Variations can then emerge because of differences not in values or interpretations of principles, but in the degrees to which IRB members as a group have fully anticipated the details of all potential logistical scenarios, foreseeing all possible on-the-ground problems.

In detecting potential logistical problems that others have overlooked in reading a protocol, IRB members cited "good catches." PIs themselves may not have foreseen all such possible difficulties. Specific details can be missed, highlighting the all-too-human aspects of processes of review—for instance, in a study's specific procedures.

A protocol randomized subjects, and then gave them the informed consent for the arm they were randomized to. It went right through the IRB. But, I said, "Did anyone think that you're randomizing people before their informed consent?" "Oh, are they doing that?" Sometimes the principles just get lost. (IRB31)

Yet in such cases, IRBs may not have "lost" principles, but simply are not fully envisioning the possible logistical details that might occur in which principles need to be applied. This interviewee admits that these inconsistencies can be troubling. "The variability is sometimes a problem," (IRB31) but not enough for her to support changing the system.

One IRB may be the only one to "pick up on" a particular problem. "We were the only IRB that found issue with that—which amazes me" (IRB13). In part, "many different eyes" viewing a protocol may reveal additional potential problems, which may be inevitable, but it is not systematic per se. However, the absence or presence at a meeting of a single individual, which happenstance could easily reverse, can also then alter the IRB's decisions, and cause inconsistencies. Differences among IRBs may thus arise due not to differences in *local* community values, but simply to *more eyes* viewing the protocol.

Defending Variations

As suggested earlier, interviewees generally defend, ignore, or downplay these variations, the subjectivity involved, and the problems that might ensue. IRBs tend to feel that their own decisions are justified. Many see themselves as "simply interpreting the regulations," and believe that their decisions are inherently valid and essentially incontrovertible, because these reflect local community values, rather than constituting subjective interpretations. Though a few chairs may acknowledge "minor" variations among IRBs in "fine-tuning and nuance," the differences in fact can be far greater. Local reviews may provide "local knowledge" (e.g., that a hospital treats vulnerable patients), but such local factual details do not generally appear to account for the variations that arose here.

Moreover, IRBs may become fixed in their stances and interpretations, developing procedures after applying the regulations in certain ways, and then rigidly forcing all protocols to comply with these procedures.

One group of PIs had 13 IRBs, and spent most of their time just trying to get through all the IRBs. The PIs pointed out that it was really minimal risk, and could be expedited. But the IRBs said, “I don’t care, this is the way we do it. It’s got to be done this way ... We don’t expedite things like this. This is the way we do things. That is our procedure—that everything like that goes to the full board.” (IRB26)

IRBs may also define “consistency” differently from each other, exacerbating problems.

We don’t disapprove studies that often. That is an example of *consistency*. But some other IRBs are a lot tougher and hard-assed than we are. (IRB40)

Yet this co-chair’s definition of consistency—“not disapproving” studies—represents a relatively low standard, and ignores the many crucial types of variations that concern not simply whether or not a protocol is approved, but *what* specific changes are requested.

DISCUSSION

These data are the first to examine how IRBs view inconsistencies among them, and how these can often result from differences not in local community values, but in personality and institutional factors, and individuals’ abilities to foresee possible logistical problems. These variations can delay or impede research, but IRBs often appear to defend these. However, there is no empirical evidence that, at least for the vast majority of research within the United States, significant systematic differences in community attitudes exist concerning the types of divergences in IRB reviews of protocols that occur. Past studies have shown that chairs define terms such as “minimal risk” differently (Bekelman, Li, and Gross 2003), but evidence has not been provided showing that these variations reflect differences in community values or local laws. Similarly, interviewees here describe having different thresholds for approving studies, struggling with whether studies are “good enough” versus should be “absolutely perfect” and “truly safe,” and making decisions about these ambiguities based on their own personal “gut feelings” and predilections, not on community values.

Though definitions of “community” can vary (e.g., as being geographic, social, or religious) (Richardson et al. 2005), 45 CFR 46, in mentioning “community attitudes,” presumably refers to the communities from which subjects are drawn, not the “community” of a particular academic department, clinic, or institution. Hence, the regulations also require that IRBs include a “nonaffiliated” member. IRBs can “ascertain the acceptability of proposed research in terms of institutional commitments and regulations, applicable law, and standards of professional conduct and practice” (U.S. Department of Health and Human Services 2009), but the variations here appear to result from subjective and idiosyncratic aspects of individual chairs, members, or IRBs, not differing institutional regulations or local laws. Moreover, IRBs even within a given institution can vary.

IRBs work hard, with very high levels of commitment, but appear to engage in highly interpretive applications of abstract principles, and in such complex human endeavors, personal factors and local institutional cultures and idiosyncrasies may be inevitable. One can argue that these individual personal predilections (e.g., being “obsessive” or having a “pet peeve”) reflect those of a larger community, but that does not generally appear to be the case. Rather, even within one institution, located in a single community, IRBs differ due to personality factors of a chair or vocal member. Reviewers may also range in their ability to anticipate and imagine possible future logistical problems in carrying out a protocol. Yet surprisingly, generally, these interviewees do not seem surprised or disturbed by these differences, and instead tend to see these variations as inevitable and even desirable, discounting the costs to research and researchers.

These data highlight how the fact that IRB members may be “self-taught” in ethics (Klitzman 2008) and, relatedly, may use “gut feelings” and the “sniff test,” not careful “ethical analysis,” poses considerable challenges in IRB review. Indeed, research in other realms has suggested that individuals frequently develop moral intuitions and feelings first, and only afterward try to determine rational ethical principles with which to articulate and justify these feelings—rather than initially engaging in ethical analysis (Haidt 2001).

These data raise questions of whether and to what degree such variations can be decreased, or are to some extent inevitable, and what the implications may be—for example, how much consistency can or should be expected, how to obtain it, and what to do if it is lacking.

As a solution to problems caused by inter-IRB variations, increased use of central IRBs (CIRBs) has been proposed for almost 20 years (Menikoff 2010; Ahmed and Nicholson 1996; U.S. Department of Health and Human Services 2011), but thus far, as of this writing, has been adopted to only limited degrees. Given logistical and other obstacles, whether such structural changes will spread, and if so, when, for which studies, to what degree, how, and with what outcomes and success, is not evident. Moreover, in most models of CIRBs, local IRBs can accept, reject, or amend CIRB recommendations. Presumably, ultimate control may thus remain local, allowing variations to continue, even if possibly to a lesser degree.

Though some critics have argued that the establishment of more centralized or regional IRBs will solve the problems caused by these variations (Green and Geiger 2006; Emanuel et al. 2004), such entities may help, but can also be subjective in their interpretations, reflecting the individual preferences of chairs or vocal members (Klitzman 2011b). Thus, regardless of whether and how much centralization proceeds, debates will doubtlessly continue, raising critical questions concerning IRB variations—their extent, causes, and possible remedies.

Other organizational changes might also be considered to help address these problems. For instance, one could imagine “licensing” PIs to do research for set periods of time, but how and whether such other approaches might work is unknown.

Another approach is to consider attitudinal changes. Specifically, IRBs should, both individually and as a group, work far more than at present to be more consistent, and value the benefits of such consistency.

The actual causes of these inconsistencies need to be far more acknowledged and accepted, not ignored and/or denied. Given that IRBs are not always systematic, it appears better to accept this phenomenon, rather than minimize it, or argue that it is not the case. Such recognition is critical to assure effective, appropriate, and just protection of human subjects. The lack of such acknowledgment causes strains between IRBs and PIs that can potentially prompt researchers not to follow IRB regulations as much as possible. IRB beliefs that variations are legitimate because they reflect different community values can foster inflexibility, antagonizing researchers, and potentially diminishing the latter's commitment to research ethics (Koocher 2005). IRBs need to recognize more, too, the costs of variations—at times unnecessarily burdening research and researchers—and to take these costs into account in reviews.

Still, attitudinal changes among IRBs will not be easy to instill. IRBs now appear to have little, if any, incentive to confront these issues or seek uniformity. In part, the costs of current variations fall on PIs, not IRBs. Each IRB may also have little desire to yield autonomy, and IRBs may fear that recognition of these subjective factors will call into question the legitimacy of their decisions. That is, if a board's review is based on personal factors, rather than community values, questions may arise as to the inherent legitimacy of the authority of that review. If an IRB's decision can be seen as idiosyncratic, it is far more open to criticism.

Some observers may argue that expanding IRB accreditation adequately addresses these issues, but accreditation focuses on standardizing IRB forms and procedures, and tends not to examine crucial aspects of the *contents* of decisions—for example, how IRBs decide how to interpret, operationalize, and apply terms such as “minimal risk” and “benefit” in specific instances, and how “obsessive” to be. Nonetheless, accreditation can potentially require that differing applications of such terms among institutions be standardized. However, the Association for the Accreditation of Human Research Protection Programs (AAHRPP) may only be able to address each institution's internal consistency, not variations that exist among institutions with different IRBs.

The development of clearer guidelines for interpreting regulations by the U.S. Office of Human Research Protections (OHRP) and the Food and Drug Administration (FDA) can also help here. The development of a body of cases and ways of resolving conflicting perspectives (as in “case law”) can assist, too, in establishing precedent. Lantos (2009) has suggested the establishment of an “appellate” IRB to allow PIs to appeal local IRB decisions at a higher administrative level. Such an “appellate” body would be publically auditable and transparent. Such an external process, perhaps through OHRP, could be helpful. Yet local IRBs may resist, rather than readily accept, such an appellate system. The lack of an appeals process may, however, further bolster IRBs in avoiding recognition of these issues. Making IRB minutes publically available (with redaction of proprietary information) can be useful as well. Within an institution, developing and establishing more of a systematic institutional memory of what was or was not approved in the past can also be beneficial.

At the same time, one might argue that investigators bear a large portion of the responsibility for the variability in IRB responses by providing insufficient information in

protocols and addressing exactly how a study satisfies regulatory requirements—for example, why it should be approved using expedited review. But IRBs may disagree with PIs' assessments, and variations have been documented in IRB responses to carefully prepared identical protocols (Greene and Geiger 2006), suggesting that inconsistencies may be due to differences in IRBs themselves.

Nonetheless, PIs may need to alter their views and approaches. Some PI complaints and perceptions of problems with IRBs may simply reflect resistance to the federal regulations themselves. The present data highlight that IRBs are usually trying to do their best, from their perspective, to protect subjects. Yet PIs may feel frustrated, and treat IRBs disrespectfully, and submit sloppy or incomplete protocols, which can exacerbate tensions. Mutual change can facilitate mutual trust, which is essential.

These issues have received little systematic attention, but are crucial to explore more fully. Further research is needed to examine among larger samples of IRBs how often and when exactly IRB variations result from institutional, psychological, community, and/or other factors, and how to best understand and address differences that may arise. Additional scholarship is needed to investigate how different groups of stakeholders, including different IRB chairs, lay and other members, administrators, university administrators, regulators, policymakers, and PIs, all view these issues and why. Definitions of “inconsistency” also vary, as do notions of how much inconsistency is acceptable versus problematic. These differences can also be further probed.

This study has several potential limitations. These data are based on in-depth interviews with individual IRB members and chairs, and did not include direct observation of IRBs making decisions, or examination of IRB written records. Future studies can observe IRBs and examine such records. Yet such additional data may be difficult to obtain if, for instance, IRBs require obtaining consent from all IRB members, as well as from the PIs and funders of protocols (to allow observation of the IRB process and examination of records). IRBs may say that their reviews of protocols are confidential, but studies of IRBs can potentially exclude identifying and proprietary information. Additionally, the interviews here examined respondents' experiences and views at present and in the recent past, but not prospectively over time to gauge whether they altered their views, and if so, why. Interviewees were not presented with hypothetical standardized protocols, but instead described actual protocols they had confronted. Use of hypothetical protocols may be limited, however, since respondents know that their responses to these are being studied, and they thus may not respond in the same way as in “real” protocols they confront. However, future investigations can use such approaches as well.

These data do not prove that variations are never due to differences in community values. However, at the very least, these interviews are valuable for shedding light on crucial areas that require further attention, suggesting several key themes and areas for future exploration. These data and their implications will no doubt meet resistance, as some IRB members or chairs may feel challenged by these findings. However, if our ultimate goals are to advance social benefits and reduce research risks as much as we can, these goals may best be achieved by understanding as clearly as possible the processes involved in IRB decision

making, and the causes of variations that arise, in order to know whether these should be addressed, and if so, how. Hopefully, these data can encourage and facilitate useful dialogue and research to help us optimally achieve these dual objectives. IRBs and researchers need to work closer together, as effectively, transparently, and equitably as they can, to protect subjects. Attitudinal and other changes, as suggested here, though difficult, are in the best interests of participants, science, and IRBs themselves.

Acknowledgments

The National Institutes of Health (NIH; R01-NG04214) and the National Library of Medicine (5-G13-LM009996-02) funded this work. I thank Meghan Sweeney, BA, and Paul Appelbaum, MD, for their assistance with the article.

References

- Ahmed AH, Nicholson KG. Delays and diversity in the practice of local research ethics committees. *Journal of Medical Ethics*. 1996; 22(5):263–266. [PubMed: 8910776]
- Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: A systematic review. *Journal of the American Medical Association*. 2003; 289(4):454–465. [PubMed: 12533125]
- Burman WJ, Reves RR, Cohn DL, Schooley RT. Breaking the camel's back: Multicenter clinical trials and local institutional review boards. *Annals of Internal Medicine*. 2001; 134(2):152–157. [PubMed: 11177319]
- Dyrbye LN, Thomas MR, Mechaber AJ, et al. Medical education research and IRB review: An analysis and comparison of the IRB review process at six institutions. *Academic Medicine*. 2007; 82(7):654–660. [PubMed: 17595560]
- Dziak K, Anderson R, Sevick MA, Weisman CS, Levine DW, Scholle SH. Variations among institutional review boards in a multisite health services research study. *Health Services Research*. 2005; 40(1):279–290. [PubMed: 15663713]
- Emanuel EJ, Wood A, Fleischman A, et al. Oversight of human participants research: Identifying problems to evaluate reform proposals. *Annals of Internal Medicine*. 2004; 141:282–291. [PubMed: 15313744]
- Geertz, C. *Interpretation of cultures: Selected essays*. New York: Basic Books; 1973.
- Goldman J, Katz M. Inconsistency and institutional review boards. *Journal of the American Medical Association*. 1982; 248(2):197–202. [PubMed: 7045428]
- Green LA, Lowery JC, Kowalski CP, Wyszewianski L. Impact of Institutional Review Board practice variation on observational health services research. *Health Research and Educational Trust*. 2006; 41(1):214–230.
- Greene SM, Geiger AM. A review finds that multicenter studies face substantial challenges but strategies exist to achieve Institutional Review Board approval. *Journal of Clinical Epidemiology*. 2006; 59:784–790. [PubMed: 16828670]
- Haidt J. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*. 2001; 108:814–834. [PubMed: 11699120]
- Jamrozik K, Kolybaba M. Are ethics committees retarding the improvement of health services in Australia? *Medical Journal of Australia*. 1999; 170:26–28. [PubMed: 10026670]
- Klitzman R. Views of the process and content of ethical reviews of HIV vaccine trials among members of US Institutional Review Boards and South African Research Ethics Committees. *Developing World Bioethics*. 2008; 8(3):207–218. [PubMed: 19046258]
- Klitzman R. Views and experiences of IRBs concerning research integrity. *Journal of Law, Medicine & Ethics*. 2011a in press.
- Klitzman R. “Members of the same club”: Challenges and decisions faced by US IRBs in identifying and managing conflicts of interest. *PLoS ONE*. 2011b in press.

- Klitzman R. How local IRBs view central IRBs in the US. *BMC Medical Ethics*. 2011c; 12(13):10.1186/1472-6939-12-13
- Klitzman R. US IRBs confronting research in the developing world. *Developing World Bioethics*. 2011d in press.
- Koocher GP. The IRB paradox: Could the protectors also encourage deceit? *Ethics & Behavior*. 2005; 15(4):339–349. [PubMed: 16578924]
- Lantos J. It is time to professionalize institutional review boards. *Archives of Pediatrics & Adolescent Medicine*. 2009; 163(12):1163–1164. [PubMed: 19996057]
- McWilliams R, Hoover-Fong J, Hamosh A, Beck S, Beaty T, Cutting G. Problematic variation in local institutional review of a multicenter genetic epidemiology study. *Journal of the American Medical Association*. 2003; 290(3):360–366. [PubMed: 12865377]
- Menikoff J. The paradoxical problem with multiple-IRB review. *New England Journal of Medicine*. 2010; 363(17):1591–1593. [PubMed: 20942660]
- Middle C, Johnson A, Petty T, Sims L, Macfarlane A. Ethics approval for a national postal survey: Recent experience. *British Medical Journal*. 1995; 311:659–660. [PubMed: 7549633]
- Newgard CD, Hui SJ, Stamps-White P, Lewis RJ. Institutional variability in a minimal risk, population-based study: Recognizing policy barriers to health services research. *Health Research and Educational Trust*. 2005; 40(4):1247–1258.
- Richardson LD, Wilets I, Ragin DF, et al. Research without consent: Community perspectives from the Community VOICES study. *Academic Emergency Medicine*. 2005; 12(11):1082–1090. [PubMed: 16264079]
- Stair TO, Reed CR, Radeos MS, Koski G, Camargo CA. and THE MARC. Variation in institutional review board responses to a standard protocol for a multicenter clinical trial. *Academic Emergency Medicine*. 2001; 8(6):636–641. [PubMed: 11388939]
- Stark AR, Tyson JE, Hibberd PL. Variation among institutional review boards in evaluating the design of a multicenter randomized trial. *Journal of Perinatology*. 2010; 30:163–169. [PubMed: 19798046]
- Strauss, A.; Corbin, J. *Basics of qualitative research—techniques and procedures for developing grounded theory*. Newbury Park, CA: Sage; 1990.
- U.S. Department of Health and Human Services. 45 CFR 46—Protection of human subjects. 2009. Available at: <http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm>
- U.S. Department of Health and Human Services. Human subjects research protections: Enhancing protections for research subjects and reducing burden, delay, and ambiguity for investigators. 2011. Available at: <http://www.ofr.gov/OFRUpload/OFRData/2011-18792PI.pdf>
- Vick CC, Finan KR, Kiefe C, Neumayer L, Hawn MT. Variation in Institutional Review processes for multisite observational study. *American Journal of Surgery*. 2005; 190:805–809. [PubMed: 16226962]

APPENDIX: SAMPLE QUESTIONS FROM SEMISTRUCTURED INTERVIEW

- How do you define research integrity (RI)? How has your IRB approached RI issues?
- Do you think IRBs differ in their views or approaches toward RI and related issues, and if so, when and how? Do you think IRBs apply standards differently regarding RI and related areas, and if so, how?
- Do you think IRBs approach conflict of interest (COI) differently? If so, how, when, and why?
- Do you think institutional factors affect your IRB's decisions about RI? If so, how?
- What do you think makes an IRB work well or not in monitoring and responding to RI?

- Do you think a centralized IRB rather than local IRBs would have advantages concerning RI and other areas? If so, what, how, and why?
- Do you have any other thoughts about these issues?

Note: Additional follow-up questions were asked, as appropriate, with each participant.

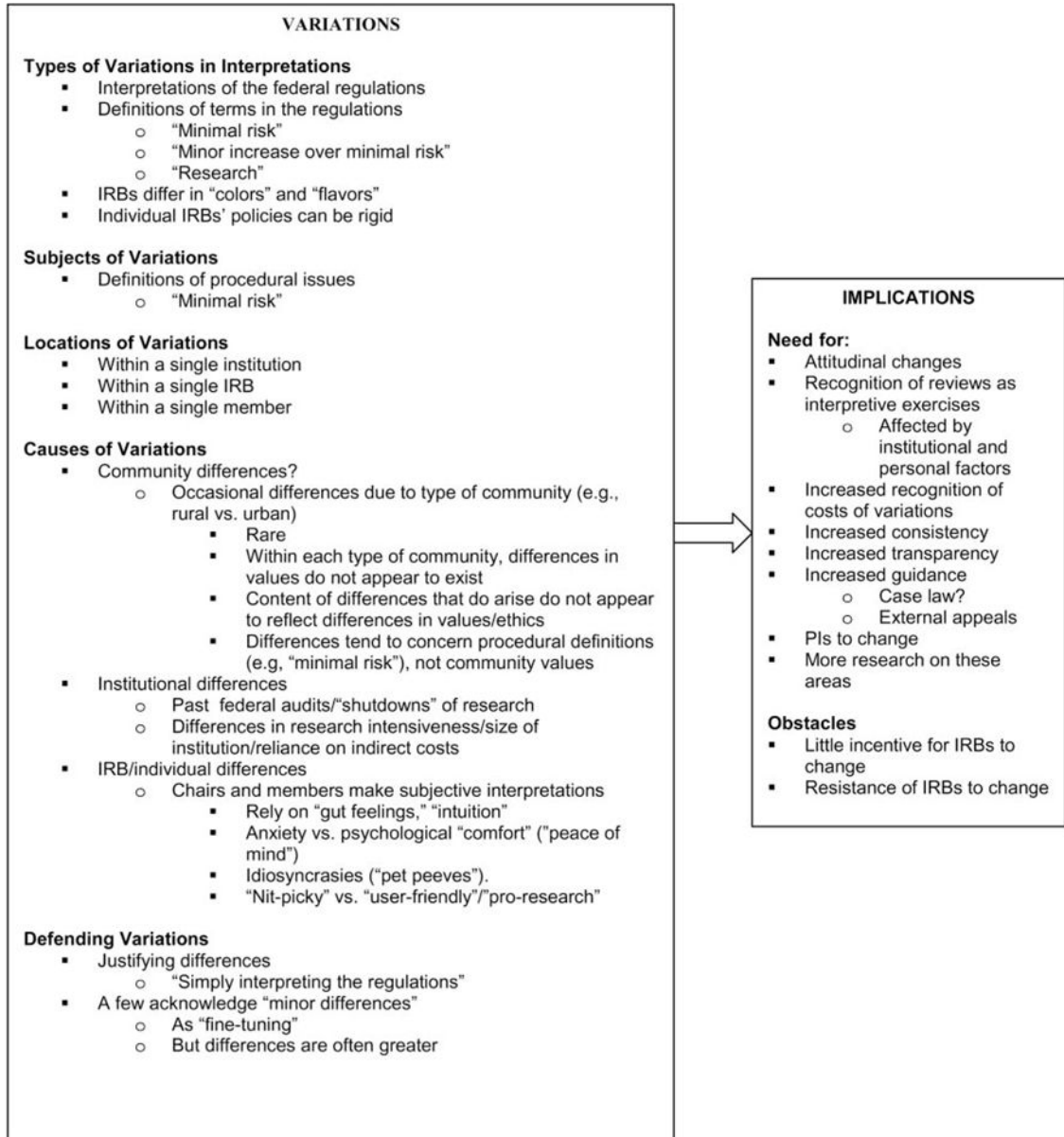


Figure 1.
Issues concerning variations among IRBs.

Table 1

Characteristics of the sample

Characteristic	Total	% (<i>n</i> = 46)
Type of IRB staff		
Chairs/co-chairs	28	60.87%
Directors	1	2.17%
Administrators	10	21.74%
Members	7	15.22%
Gender		
Male	27	58.70%
Female	19	41.30%
Institution rank		
1–50	13	28.26%
51–100	13	28.26%
101–150	7	15.22%
151–200	1	2.17%
201–250	12	26.09%
State vs. private		
State	19	41.30%
Private	27	58.70%
Region		
Northeast	21	45.65%
Midwest	6	13.04%
West	13	28.26%
South	6	13.04%
Total number of institutions represented	34	

Note. From Klitzman (2011c).