The IRB Sledge-Hammer, Freedom and Big-Data

Curtis Naser, Ph.D.

Associate Professor of Philosophy & Applied Ethics

Fairfield University

cnaser@fairfield.edu

203-414-1132

Curt Naser is a bioethicist and has served on biomedical and social science IRBs for over 20 years. In addition Curt spent over ten years coding academic data management systems. He is the co-founder of Axiom Education, a company that provides IRB management software that is in use in universities, hospitals and research institutes around the country.

The application of the regulatory model of IRBs to research using big-data, particularly in the for-profit sector, raises a host of issues. I want to highlight three areas of concern where the regulatory framework of IRBs is challenged by this research: the power of the IRB, privacy, informed consent and the obligation of subjects to participate in research.

The practice of ethical and regulatory review of research involving human subjects has its roots in revelations of some very harmful research that came to light in the early 1970s. The Tuskeegee Syphilis Study is perhaps the most notable, in which a cohort of African American men from Alabama were denied treatment for syphilis with simple antibiotics over a period of over 20 years, resulting in many untimely deaths and the transmission of syphilis to their wives and children. Other studies that came to light involved infecting mental patients with hepatitis and injecting radioactive agents, all without consent, much less informed consent. The public was primed by the civil rights movement and the culture of the day to demand change and the protection of vulnerable patients. However, medicine had made enormous strides in the development of life-saving, miraculous treatments and cures: vaccines, antibiotics and mechanical ventilation to name just a few. Congress had to act, but recognizing the value of biomedical research, and most importantly perhaps, its own lack of expertise, Congress created a expert “Commission” whose responsibility it was to make recommendations to what was then HEW (now HHS) that HEW was required to implement *without* any further approval by Congress. This was an extraordinary moment when Congress recognized its own ignorance and entrusted a very sensitive issue to a panel of experts.

The result of the work of the “National Commission for the Protection of Human Subjects in Biomedical and Behavioral Research” (or just the “National Commission”) was the federal regulations which we have in place today: 45CFR Part 46.[[1]](#footnote-1) These regulations have served us well in the domain of biomedical research and though they are not without their problems, they have had the effect of creating a culture of ethical awareness around research involving human subjects, which is perhaps as important as the actual IRB review process that they engendered.

The IRB’s Sledge Hammer

Recognizing that inserting an ethical review process between researchers and their human subjects can lead to conflict with both the researcher and their institution, the regulations set out a very unique requirement for IRBs that is simply unheard of in almost every other organizational and political context:

“Research covered by this policy that has been approved by an IRB may be subject to further appropriate review and approval or disapproval by officials of the institution. However, those officials may not approve the research if it has not been approved by an IRB.” (45CFR46.112)

No one can overturn an IRB’s decision to not approve research. I like to refer to this as the IRBs sledge hammer and it accounts for both the success of IRBs as well as some of the problems that IRBs have given rise to.

On the success side, this authority makes it very difficult for a researcher, their department or the institution itself to influence decisions by the IRB. As federal research dollars are at stake in the conformity of research to the regulations, no institution is going risk those millions of research dollars by attempting to influence an IRB’s decision. You will find, however, that few research studies are actually not approved by IRBs. Instead, they are sent back to the investigators with suggestions for changes and clarifications. The investigators are motivated to do what the IRB demands because they have no alternative if they wish to do their research. As a result, serious institutional and interpersonal conflicts of interest are negated and IRBs are largely free to apply their judgments and ethical reflections. (This is not to say that individual IRB members have never been lobbied by their colleagues and superiors, but thus influencing a whole committee, which is also required to have an unaffiliated member, is far less likely.) In most cases, however, and as the National Commission envisioned the IRBs that their recommendations mandated, the review process is collegial and provides a layer of ethical and regulatory review that is helpful to investigators. It can and often does result in better science and better recruitment of subjects.

On the downside of the IRBs sledge hammer is the problem of mission creep. Because few people in most institutions that have IRBs actually understand the regulations and how IRBs operate (including sometimes the IRB members themselves!), and because researchers must simply submit to the demands of the IRB to get their research approved, some IRBs over time have extended the scope of their authority beyond what the regulations require and without the explicit authorization of their institutions. For instance, IRBs are explicitly prohibited by the regulations from considering the social or political impact of research when reviewing it. Yet some IRBs will do just that. A more common creep in IRB mission is to meddle in the scientific design of studies when such studies represent little or no risk to the human subjects.

So the bottom line is that the IRB sledge hammer of un-appealable decisions to not approve research ensures independence of judgment by the IRB, but at the risk of allowing the mission of the IRB to creep into areas it is not specifically charged to exercise such a great authority. Since the hazards of institutional coercion are great and the hazard of mission creep is remediable through education and good governance practices, the benefits of making IRB decisions un-appealable outweighs the hazards and serves to protect human subjects.

If IRBs are defined in part by this un-appealable authority, otherwise they become merely advisory, the question is whether this model is a good fit for corporate and private organization research that involves human subjects. The answer to that question will depend upon how committed the organization is to ensuring the protection of human subjects and how much of their own independence these organizations are willing to give up when it comes to this kind of research. One should also note that with academic IRBs and biomedical IRBs, the members have more autonomy of judgment due to the cultures of these institutions. In the university, tenured and tenure track professors populate IRBs. Academics are notoriously independent and are much less open to institutional persuasion and coercion. Biomedical IRBs are populated with clinicians who also enjoy a certain amount of autonomy within their organizations. With corporate IRBs, the personnel may not have the same type of independence and would thus be more at risk of influence by managers and other superiors.

Can a private entity tolerate the creation of an organizational unit that has veto power over its own research projects? This is a big step for any business or organization to take. Universities and hospitals, pharmaceutical and device manufacturers and other research organizations funded by the federal government simply had no choice but to accept the regime of IRB review and all that it demands. Certainly the private sector can engender more of the public’s trust by going in this direction, and many corporations and private organizations have statements of ethical principles and policies. But for IRBs to operate *ethically*, they need the independence that the regulations have granted them in those institutions under federal oversight.

Privacy

The purpose of the IRB is to protect human subjects in research. What that means is that the research presents an at least an even if not favorable balance of risks and benefits to the subject, and that the risk to subjects is compensated by the value of the knowledge that results. In the biomedical domain, risks and benefits are relatively straightforward, provided you have the expertise to fully understand them. There are also risks to subjects related to a breach of confidentiality, but these risks are treated separately in the federal regulations. That is, confidentiality is an expectation and researchers are required to protect the privacy of the data they collect, but the risk of breach of confidentiality is not treated as a “research risk” but instead as a risk that can be managed by good practices to protect data.

This is a vital distinction to make because in social science research as well as research involving the analysis of “big data”, the primary risk to subjects is precisely the breach of confidentiality. The federal regulations, written at a time when the possibilities of “big data” were not foreseen, goes so far as to consider research “minimal risk” that generates sensitive data provided that adequate measures are taken to protect the confidentiality of that data. In a policy statement made in 1998 revising the list of categories under which protocols may receive “expedited review”, OHRP wrote that the expedited review procedure, which can only be invoked for research determined to be minimal risk…

“…may not be used where identification of the subjects and/or their responses would reasonably place them at risk of criminal or civil liability or be damaging to the subjects financial standing, employability, insurability, reputation, or be stigmatizing, *unless* reasonable and appropriate protections will be implemented so that risks related to invasion of privacy and breach of confidentiality are no greater than minimal.”[[2]](#footnote-2) (emphasis added)

I do not believe that the regulators conceived the possibilities of how much data can be aggregated on individuals and how powerful the analyses of that data can be in drawing conclusions about a person’s beliefs, practices, habits, behaviors, etc. In a parallel instance, the regulations also never anticipated the power of genomic analysis. The regulations, for instance, specifically exempted from IRB review use of medical tissue specimens if those specimens already are in existence (e.g., stored in a hospital pathology lab or blood bank), or likewise the review of existing medical records, in both cases provided that the investigator does not record any identifiers or links to identifiers in their data. The problem, however, is that genomic information is a unique identifier all by itself and contains a great deal of information about the person from whom it was taken. With the inclusion of genomic information in medical records, the anonymity of tissue specimens is at best suspect and at worst now, a fiction. Likewise, abstracts of medical records may contain sufficient granularity of data that even without explicit identifiers, the data can easily be matched back to those identifiers through data query techniques.

On more than several occasions, investigators using anonymous tissue specimens have discovered that the patients from whom the specimens were derived were at risk of serious and debilitating disease. With genuine anonymity, there can be no obligation to warn these patients. But what obligations do the researchers have when they can in fact decode the identity of any sample they have in their position? On the “big data” side, imagine that researchers are able to correlate consumer data with medical information and come to learn that certain patterns of consumer buying indicate an undiagnosed disease, a propensity to child abuse or other violence. Would the researchers have an obligation to inform the person or law enforcement authorities? My point is not to specifically answer the ethics of that question but to demonstrate that the practice of research is such that it will lead to unexpected findings and place the researcher into difficult ethical quandaries. It is best to prepare in advance for them. Given the power of current computational techniques, it is unlikely that any “big data set” can genuinely be anonymous, even if there are no specific identifiers attached to it.

Informed Consent

The Belmont Report[[3]](#footnote-3), issued by the National Commission and laying out the ethical framework for research involving human subjects, sets out the principle of respect for persons as the ethical basis of informed consent. The concept has its origins in the moral philosophy of Immanuel Kant and is often referred to by Kant’s term: *autonomy*. It articulates the moral obligation to respect the freedom of others to engage in activities of their own choosing and to not obstruct them in achieving their own goals and actions. Informed consent is most commonly derived from this moral principle, though informed consent also serves to protect human subjects from harm.

Informed consent is an ideal that is often difficult to realize in practice. Certainly in the domain of biomedical research, while the researcher may attempt to inform the subject of the nature of the research, the subject’s background, educational level and state of health may all work against any full, much less partial understanding of what they are consenting to. They are afforded the choice to participate and researchers make the effort to educate them but studies have shown that subjects often misunderstand even that they are in a research study. Culture and beliefs can also pose a barrier. Imagine trying to consent a subject into an HIV study who does not have any grasp of the germ theory of disease or what viruses are.

We all have been offered and most of us carry barcoded tags from our local grocery store, drug store and other consumer outlets. These barcodes are often the condition of getting discount prices. They are used by the company to build profiles of their customers which is used for targeted advertising and which may be sold to other consumer research companies. Few of us read the fine print in the agreements that we commit to when we get these cards. Consumers have very little understanding of how their purchase data is used, but we are goaded/enticed/coerced into signing up to get those special sale prices.

The internet has grown up on a model of small coercions: we are given free stuff like search engines, social media, software, etc. In return, the companies that provide these services collect and use our data, Google Mail going so far as to scan our emails to learn what products to advertise to us. If we were asked explicitly: can we read all your email so that we can develop a marketing profile of you, few of us would consent to such practices. But millions sign up for Google Mail because it is free and easy.

It is these market driven practices that are the source of a great deal of data that is used in corporate research. There are, of course, other sources of data that may be garnered: credit and financial data, real property ownership data, toll collection data, medical records, video face recognition data, and so on. Much of this data was provided for a single purpose, yet it can be aggregated and used in many ways and for many other purposes, some to which we might agree and others to which we would never agree. This suggests that the provenance of the data that is the basis of so much “big data” analysis is suspect. That is, the data was not explicitly provided to those who would analyze it for the purposes of those analyses.

An interesting parallel can be drawn to this situation with embryonic stem cell research. There are IRB like committees (ESCRO – Embryonic Stem Cell Research Oversite committees) that review research involving embryonic stem cells. It is vital to the researchers to ensure that if they are developing novel treatments that they use stem cells for which consent was explicitly given. Were they to use stem cells for which there is not clear provenance of informed consent, their work will not be commercially viable and they will have to reproduce it using properly consented stem cells. One might argue that the “parents” are not in the least bit affected by the use of these cells, but they are recognized as having an interest in the disposition of their embryo’s cells. The situation is structurally not much different with the ways that personal data is used, manipulated, bought and sold and ultimately comes back to us in the form of targeted marketing. In both cases, information that is “about us” is used for purposes that we never envisioned, sanctioned, understood or consented to.

Indeed, the ability to aggregate data from diverse sources allows researchers to paint very detailed representations of persons’ daily lives. It is one thing to give up some of my personal information for some discounts, but none of us have knowingly consented to such pervasive accumulation of data. As people often remark, it gives you the creeps when your computer starts showing you ads for products you recently purchased at the hardware store. If the social construction of identity is a dialogue between the narrative of our own self-construction and the multiple narratives of ourselves offered by others, the power of big data to construct a master narrative of our lives threatens the freedom of self-identity. We can be nudged and manipulated by the echo chamber of our past reflected back to us in the news, information and advertising targeted to us through data algorithms working at the behest of unknown corporate entities. The European “right to be forgotten” may be an attempt to give individuals control over the dissemination of their past across the internet, but it points to a much deeper notion that our freedom of self-identity is threatened by the accumulation of widely dispersed data about us.

The Obligation to Participate in Research

At its root, the problem of big data is one of control. Should the persons from whom data is derived have any say in how that data is used? To accept that people should have such control can create conditions that will prevent a great deal of research from moving forward. In the medical context, requiring subjects to consent to the use of their tissue specimens or medical records in research would stifle a great deal of basic genetic and epidemiological investigation, studies which are vital to the advancement of medical science and medical practice. Indeed, we are all patients in a health care system built upon research performed on prior patients’ records and specimens. We are the beneficiaries of that research and one can plausibly make an argument that we as patients all have an obligation to contribute our own data and specimens to benefit future patients. There is also the practical matter of getting consent from patients long gone from the hospital to use their specimens and records.

We can see a parallel with the non-profit sector. Imagine an NGO that provides disaster assistance. They may find it invaluable to collect data on their services and the individuals who receive those services so that they can improve their practices and provide better service to victims in the future. Thus, as a current victim, I have benefitted from the research on past victims and have some obligation to contribute to the improvement of services to future victims. Few beneficiaries of disaster relief services would have any objections. I am not so sure that in the for profit domain this idea of serving to help future populations holds quite as much moral sway. One would have to look at it on a case by case basis. For instance, a firm that uses my financial and consumer data to target advertising to me is providing to me little or no benefit, but is providing its customers to whom it sells this data great benefit. The advertising is designed to divert my desire and part me from my hard earned cash. On the other hand, an automobile manufacturer that gathers real time data on performance of its vehicles may be using that data to improve future vehicle safety, emissions, etc. So the principle that we look at what obligations subjects might be fulfilling through their unconsented (or poorly consented) provision of data about them would be a good yard stick to measure the propriety of the use of that data.

IRBs are designed to protect human subjects in research. They need independence to freely make judgments but those judgments need to be constrained by an agreed upon framework, as is currently provided by the federal regulations. These regulations provide a framework for consideration of risks and benefits, but the harms to subjects will include not only the traditional concerns about breach of confidentiality in this type of research, but a more fundamental concern about the relationship between information and human freedom.

1. 56 FR 28012 (June 18, 1981), http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html [↑](#footnote-ref-1)
2. [63 FR 60364-60367](http://www.hhs.gov/ohrp/policy/63fr60364.html) (November 9, 1998), http://www.hhs.gov/ohrp/policy/expedited98.html. [↑](#footnote-ref-2)
3. The National Commission for the Protection of Human Subjects in Biomedical and Behavioral Research, *The Belmont Report*, (April 18, 1979), http://www.hhs.gov/ohrp/policy/belmont.html. [↑](#footnote-ref-3)