



The Ethics of Data Access, Use, and Sharing for Human Subjects Research

March 1, 2017

Boston, Massachusetts

Workshop Summary

Table of Contents

- [Executive Summary](#)
- [Workshop Summary](#)
- [Part I: Perspectives on the value of research employing new approaches to the access, use, and sharing of personal information](#)
- [Part II: Conceptualizing an individual's rights, interests, and expectations with regard to their health and other data](#)
- [Part III: Implications for consent, authorization, and data stewardship](#)
- [Part IV: Responsibilities for ethical oversight of research involving data access, use, and sharing](#)
- [Conclusion: Summary and next steps](#)

Executive Summary

On March 1, 2017, PRIM&R hosted a workshop entitled the Ethics of Data Access, Use, and Sharing for Human Subjects Research.

The workshop convened patient advocates, researchers, institutional representatives including institutional review board members, medical journal editors, ethicists, and policymakers to examine how the collection, access, sharing, and use of an individual's data for research purposes across a broad range of contexts presents new scientific opportunities and raises novel ethical challenges. The overarching goal for the workshop was to examine the question of whether our current ethical and regulatory framework adequately protects the rights and interests of individuals and their personal information in the setting of rapidly evolving research policy and practice.

PRIM&R's [Public Policy Committee](#) selected four ethically complex areas for participants to consider, including 1. the value of research with shared data; 2. individuals' rights and expectations with regards to their data; 3. consent, authorization, and data stewardship; and, 4. responsibility for the ethical oversight of research involving data access, use, and sharing.

Participants concluded that, though research with shared data is likely to significantly advance science and the state of health care, there are a number of ethical barriers which must be addressed to fully realize the value of data sharing. The following emerged as broad recommendations for how the research oversight community might begin to address these barriers:

- Identify a set of common principles for research with shared data which addresses: community engagement; making research relevant to communities; improved education of key stakeholders; honoring people's desire to be asked regarding collection and use of their data; and clearer, more consistent definitions of terminology for those working with shared research data.
- Identify appropriate institutional bodies, apart from or in addition to the IRB, to oversee policies and practice related to data access and use.
- Undertake efforts to enhance transparency and improve accountability in research with shared data in order to build trust with those who are being asked to make their data available.
- Create new mechanisms to secure the buy-in and authorization of the participants whose data will be shared, beyond traditional consent, that assure them that their interests in their data are honored and the uses to which data will be put are truly in the common good.
- Identify non-institutionally-based approaches for addressing institutional conflicts of interest around data ownership and access in order to re-build public trust in research.
- Improve investigator and institutional education on the value of research with shared data, but also on the requirements and principles of data ownership, stewardship, and good data security.
- Revisit and revise funding and other incentives to encourage sharing and use of shared data among researchers, and identify mechanisms for improving trust among researchers who are working on shared data initiatives.
- Address the current logistical challenges of using shared data including interoperability concerns, data quality and relevance issues, and non-standardized data curation practices.

PRIM&R looks forward to working with interested stakeholders on implementation of some of these recommendations.

Additional information on the workshop, including participant biographies and a comprehensive bibliography, is available on [PRIM&R's website](#).

Workshop Summary

The workshop convened patient advocates, researchers, institutional representatives including institutional review board members, medical journal editors, ethicists, and policymakers to examine how the collection, access, sharing, and use of an individual's data for research purposes across a broad range of contexts presents new scientific opportunities and raises novel ethical challenges. The overarching goal for the workshop was to examine the question of whether our current ethical and regulatory framework adequately protects the rights and interests of individuals and their personal information in the setting of rapidly evolving research policy and practice.

The workshop was organized along four main topic areas. Each part began with a series of brief presentations by invited panelists followed by a moderated roundtable discussion which further explored the topic and themes emerging during the presentations.

After welcoming participants to the workshop, PRIM&R's executive director, Elisa A. Hurley, PhD, summarized the central questions and structure of the day and led introductions of all roundtable participants.

Part I: Perspectives on the value of research employing new approaches to the access, use, and sharing of personal information for research

The moderator, David H. Strauss, MD, of the New York State Psychiatric Institute and Columbia University's Department of Psychiatry, introduced Part I by noting that because new approaches to data access and sharing force us to rethink the ethical frameworks that underlie human research ethics, it makes sense to first examine the scientific value and the promise associated with such initiatives.

[Laura Lyman Rodriguez](#), PhD, of the National Human Genome Research Institute of the National Institutes of Health (NIH), discussed the multiple ways the NIH is seeking to improve the diversity of the population whose data it collects, and using shared genomic data to give researchers the opportunity to validate their own data and advance public health and the care of individual patients. She discussed how the following projects are all changing how researchers combine and review data:

- The ENCODE Project: ENCyclopedia Of DNA Elements
- Genotype-Tissue Expression Project (GTEx)
- ExAC Browser (Beta) Exome Aggregation Consortium
- eMERGE network

Speaking on behalf of the *British Medical Journal*, a member of the International Committee of Medical Journal Editors (ICMJE) Elizabeth Loder, MD, MPH, of Harvard Medical School, indicated that medical journal editors believe that data sharing allows researchers to check the accuracy of published research, ensures that data collected is put to its best use, and improves the quality of research. However, she emphasized, the journals have limited oversight functions and enforcement capabilities.

[Sally Okun](#), RN, MMHS, from PatientsLikeMe, argued that patients overwhelmingly want to share their data to improve future care and research, citing an Institute of Medicine study showing that 94% of the Americans they surveyed were willing to share their data to improve

care and help patients with similar conditions.¹ PatientsLikeMe recognizes the importance of the patient voice and works to collect patient input and identify ways to “give back” to patients.

[Irene Pasquetto](#), MA, from the Center for Knowledge Infrastructures, UCLA, discussed her work showing that researchers are for the most part not re-using data that is available in open repositories prior to publication. The reasons for this include non-standardized data curation practices, data quality concerns, trust issues, and lack of funding to help researchers actually reuse the now open data sets. She raised the question of whether sharing on a per-request basis might be more cost effective than sharing in open repositories.

[Joseph Ross](#), MD, MHS, from Yale University, discussed how his concerns about selective reporting and publication of research that impacts the quality of clinical care prompted him and colleagues to create the Yale Open Data Access (YODA) Project, a data sharing platform that works with industry and other partners to promote open data, support the responsible sharing of clinical trial data, and address concerns about cost and effort of sharing data.

Nancy Kass, ScD, from The Johns Hopkins University, noted that, despite billions of dollars being spent on research and millions of people volunteering to participate in research, a great deal of healthcare is delivered without an evidence base. Dr. Kass is involved in the learning healthcare system model and the All of Us research program, two innovative attempts to expand the evidence base for healthcare delivery. She also underscored what she believes is a mistaken tendency to frame the advancement of science and individual interests as in tension, suggesting that this view is a symptom of not doing our science right—when we *are* doing it right, the collective interest in advancing science and individuals’ interests in the uses of their data are aligned. At the same time, she reminded the roundtable that to say that data sharing has ethical promise is not to say that it is in and of itself an ethical good.

Adrian Hernandez, MD, MHS, FAHA, from Duke University, spoke about PCORI’s National Patient-Centered Clinical Research Network (PCORnet) and NIH’s Health System Collaboratory, both of which are working to unlock the value of the enormous amount of data that is generated in the health care system. He made clear that the will and the technology to make data available is there, but there continue to be structural barriers, including willingness to enter data use agreements. Furthermore, some health systems are reluctant to participate in data sharing because of proprietary concerns.

Dr. Strauss then asked the panelists to comment on what steps the research community needs to take to unlock the value in shared data. The following themes emerged in the course of discussion: The need for transparency about data sharing and the goals of science; the importance of public and stakeholder engagement, along with transparency, for building trust; the role of institutional and scientific culture; and the need to address structural, practical, and governance concerns.

Dr. Hernandez and Ms. Okun agreed that one significant problem is we are not as effectively engaging patients around the collection and sharing of their data as we could be. Research suggests that, if we had in place good systems for regularly communicating about data sharing and why it is important during routine encounters in the clinical health care system, patients would for the most part willingly share their data. In fact, Ms. Okun mentioned studies showing that patients are surprised when they learn that their data are not already being shared to

¹ Grajales, F., D. Clifford, P. Loupos, S. Okun, S. Quattrone, M. Simon, P. Wicks, and D. Henderson. 2014. [Social networking sites and the continuously learning health system: A survey. Discussion Paper, Institute of Medicine](#), Washington, DC.

improve public health. Making the case routinely and regularly can increase transparency and build trust, but we have not been very good at doing that.

A question was raised whether the promise of regular patient engagement can be generalized to communities that feel disenfranchised from the biomedical establishment and the goods of health care. Ms. Okun responded that it is possible to engage such communities in culturally sensitive ways that are transparent about the reasons that a researcher might want data from a particular community (namely to diversify a research study), that engage the community as important partners from whom researchers can learn about the problem being studied, and that ensure that data collected appropriately is going to be valuable to them and their needs. The group also talked about using technology to our advantage to be able to tailor our engagement of individuals to their wishes.

Dr. Kass made the point that patients are not the only stakeholders who have a crucial role in the value proposition of shared data and whom we need to do a better job engaging. She hypothesized that one central reason there has not been greater participation in the learning health care system model has to do with institutional culture, specifically, institutional leaders failing to believe in the model, commit to changes required, and buy into the value of ongoing data collection and sharing. The same barriers had to be addressed when quality improvement initiatives first began.

Also on the issue of culture, Ms. Pasquetto argued that it matters how we frame the value of data sharing; for example, researchers are more willing to share their data when they are told the data are valuable and could be reused than when sharing is framed in terms of the need to replicate research results. This has implications for the funding of data sharing projects and incentivizing people to participate in such projects.

Closely related to the barrier of institutional culture and buy-in from institutional leaders is the issue of structural barriers to widespread data sharing. Workshop participants pointed out that data sharing is often an “add on” requirement to research, coming from the outside (e.g., journal or funder mandates). As a result, data sharing feels burdensome, and it is not surprising that it has been hard to get buy in on the ground. Dr. Kass argued that this might be addressed if both clinicians at the point of care and patients are engaged, from the start, in the design and structure of research in which sharing is emphasized. Dr. Hernandez mentioned that when they speak to participants and clinicians about the value of sharing data they emphasize return of results and improvements in the quality of care, which are issues that patients and clinicians care about.

Dr. Grady noted that designing research with data sharing in mind will help not only with research culture barriers but also allow researchers to develop data that is more user friendly for future researchers. Mr. Wilbanks provided examples of how Sage Bionetworks has successfully encouraged large companies to share their data by working with them at the start of the research process to identify where their companies might see mutual rewards by sharing their data.

Workshop participants also discussed practical barriers to widespread access and use of data, citing how hard it is to work with data that has been shared. Much of this has to do with the lack of interoperability² of the platforms and the relevance and quality of the data shared. The point

² Where “systems can exchange and use electronic health information without special effort on the part of the user.” The Office of the National Coordinator for Health Information Technology. [Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap](#), Washington, DC.

was made that if the systems are usable, relevant, and clean, then a cultural shift will be easier to “sell” to researchers.

Finally, workshop participants also discussed challenges around funding and governance. Many of the workshop participants emphasized that investigators need time and financial resources to work on data sharing initiatives so they do not become “unfunded mandates.” There was also agreement that a system of common governance would make it easier for researchers and institutions to take part in new data access and sharing initiatives. Any new governance structure should address practical concerns such as data usability and be sensitive to the need for different approaches for the different kinds of shared data and access initiatives.

Part II: Conceptualizing an individual’s rights, interests, and expectations with regard to their health and other data

The moderator, Leonard Glantz, JD, of Boston University School of Medicine, introduced Part II by noting that the research community is still working to determine whether and how medical data differs from other kinds of research data, and that there are different views about what privacy refers to. It is important to be clear about what we mean by “privacy” when conceptualizing someone’s interests in his or her data.

[Alan Rubel](#), MA, JD, PhD, from the University of Wisconsin-Madison, outlined different ways to conceptualize the values that underlie privacy rights, including consequentialist conceptions (i.e., balance the aggregate good that can come from data sharing against individual harms from potential privacy loss), and autonomy (also eudaemonist) conceptions (i.e., people have some interest and ability to decide what matters to them and act according to those decisions), and he raised questions about whether it is possible to act autonomously in the realm of data sharing, given people’s limited understanding of how their information is shared and used. He also discussed the issue of fairness of distribution, noting that, especially with respect to commercial sharing of data, the people who benefit are often not identical to the people who contributed their data.

[Suzanne M. Rivera](#), PhD, from Case Western Reserve University, argued that “research exceptionalism” is problematic and unjustified because individuals increasingly share their private (and even health) information on social media and with commercial entities that are not directly tied to research, IRBs already evaluate the use of research data for potential risks to human subjects, “respect for persons is not the only Belmont principle,” and subjects can be protected if researchers are held accountable for unauthorized behavior. She suggested we need to change our thinking from a focus on individual rights, to a focus on the common good.

Celia Fisher, PhD, from Fordham University, focused on concerns related to socially marginalized populations and group harms. Typically, decisions about risk and benefit have been made by IRBs, who share an assumption that scientific progress is a social good, but that perspective is not always shared by those with less power and influence who may be most affected by future health policies informed by big data research. Using examples, she questioned whether the evaluation of research risks and benefits sufficiently takes into account how scientific data can be used to sustain social stigmatization and discriminatory health policies, especially when much of the research is funded by private sector and government entities who have interests different from those of minority groups, and when the subjects involved and local IRB exert less influence over the downstream uses of the data. We need innovative approaches that foster public transparency and stakeholder participation in the risk-benefit analysis of how big data are used.

Michael Zimmer, PhD, from the University of Wisconsin-Milwaukee, presented three “provocations” to the group. First, he suggested that when we examine the use of large data sets for medical research through the lens of information ethics, we are better able to understand and potentially address widely held views about consent, anonymity, and privacy that seem irrational or inexplicable when we look through the lens of regulatory compliance. Second, he proposed that a conception of privacy as “contextual integrity” provides a useful way to analyze individuals' rights and expectations regarding who will have access to their information (e.g., people don't expect information they share on Facebook to become part of a research dataset, though they do expect it to be available to their friends). Third, he urged the group to recognize that data are not homogenous, and that there are various considerable complexities in thinking about the initial and secondary uses of different types of data.

Ifeoma Ajunwa, JD, PhD, from the Berkman Klein Center at Harvard University, proposed that we ask whether the term “human subject” should include not just someone who participates in a clinical trial, but also “customers” of a service such as 23andMe, which has removed genetic testing from the clinical setting and made it a direct-to-consumer service. Both 23andMe and workplace wellness programs are selling their customers' and employees' genetic information for research purposes, and she questions whether customers and employees understand that their data are being sold to undisclosed parties who may not always be acting for altruistic research purposes. She asked what protections are available to subjects in these new research spheres, and whether private companies should be allowed to determine who is a subject, what data may be shared and used, and how.

Mark Barnes, JD, from the Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard University (MRCT Center), spoke about the new reality of data use, in which commercial firms are increasingly trying to buy both de-identified and identifiable forms of information, including a person's financial data, consumer data, biospecimens, and phenotypic data, and working to capitalize on the aggregation of these different types of data. He also noted a surprising provision in the new Common Rule, namely, that if someone revokes his/her broad consent for the future use of his/her identifiable data or biospecimens, institutions are permitted to deidentify that data or biospecimens and keep using it for research. This reflects a clear “value choice.” He concluded by suggesting that this new era of data sharing should be accompanied by strong enforcement mechanisms targeted at parties who re-identify data or use data in ways that were not authorized.

Elizabeth Buchanan, PhD, from the University of Wisconsin-Stout, asked what values and principles underlie our decisions about how to share and use data. She believes we need a human rights framework for thinking about data sharing. When we look at who benefits and who stands to lose from data sharing efforts, the playing field is not fair. The issue of data sharing is often framed as one of individual rights versus societal benefit. If we come down on the side of societal benefit, we might ask whether individuals should have the right to opt out of pervasive data collection. But this is only reasonable to ask against a background of shared values and principles around data sharing, which we do not currently have. There are different types of data and different controllers of data. For instance, commercial entities like Facebook do not always make it known that they will sell or share data, or for what (commercial) purpose. Entities such as PatientsLikeMe, in contrast, are transparent about the fact that data will be shared, and are interested in data sharing for altruistic reasons. We need to be concerned about how we are educating the next generation of data scientists about the values and ethos of big data and data sharing.

Professor Glantz asked the panelists if we could mitigate the potential harms that arise from data sharing by keeping private information out of the hands of bad actors and those who seek to share data only for financial gain, and whether there is any harm in collecting de-identified private information and providing it to “good” actors who just want to use the data for altruistic purposes.

The following themes emerged in the course of discussion: The need to re-think and re-define harm and privacy violations in an era of data sharing; concerns about how economic benefits are distributed under emerging data sharing models; the importance of reviewing who decides what “public interest” means when data are shared, and the need to ensure that people understand the ways in which they are relinquishing traditional rights to privacy.

Dr. Rubel suggested that even when participants are not harmed by new data sharing and access initiatives and stand to benefit in general from scientific advances, they may not always see the same financial benefits as the organization who uses their data. This raises concerns about fair distribution of benefits. Workshop participants discussed how the Henrietta Lacks case brought questions about the fair distribution of economic benefits to the public’s attention, and whether people would have the same concerns had her cells remained de-identified.

Dr. Zimmer urged the workshop participants to go beyond the traditional notions of harm such as economic consequences and look at the implications of data sharing for autonomy and an individual’s ability to decide how his or her data are used even after de-identification measures have been put in place.

Dr. Ajunwa agreed that loss of autonomy is a relevant harm here, and raised the point that an individual’s ability to control his or her health data is directly tied to his or her personhood. She noted that individuals are increasingly asked to give up their privacy rights in the name of “progress” or “innovation,” when it is really in the service of surveillance; we should think beyond economic harms and think instead in terms of people’s dignitary rights. Dr. Rivera argued that we should compare the risks and benefits of increased data access and sharing, and that in some situations, there may only be a small risk of theoretical dignitary harm and a great scientific reward. However, she also pointed out that trust in the scientific enterprise is important, and that there should be penalties for bad actors especially for individual investigators who misbehave. Mr. Barnes suggested that IRBs need a better conceptual framework to identify both economic and non-economic harms “discrete and insular minorities” might face, such as religious and dignitary harms.

Mr. Barnes also floated the “free rider problem,” whereby if the majority of the public chooses to opt in to a system in which their data are shared, the minority who opt out benefits from advances in science that result, and faces no potential dignitary harms themselves. Dr. Rubel pointed out that even though individuals might choose not to participate in data sharing or research, they may support the scientific enterprise in other ways, including with their tax dollars.

Dr. Fisher suggested that we should not assume that even publically funded research is always in the interest of the public good. First she questioned who this “public” is. Second, we need more transparency on the part of funders, including the government, about the purposes of research they are supporting. Using personal biomedical data for future research on social issues (for example, the NIH Violence Initiative, which looked at potential biological causes of crime in urban environments) has the potential to lead to group harms such as stigmatization or problems getting insurance. There has to be more oversight with respect to future harms associated with use of our data that we can't anticipate now. Such concerns might be mitigated

by public transparency and soliciting stakeholder input, as both measures influence what research the government chooses to fund.

Sharon Shriver, PhD, from PRIM&R, raised the concern that much of the discussion during the workshop assumes that research data can be fully de-identified. However, our ability to de-identify data (especially genetic data) might be constrained by future technological developments. Workshop participants discussed how the possibility of criminal penalties for re-identification of de-identified material might deter such practices in the future. Other participants raised the problem that the more data are de-identified, the less useful it is for research purposes.

Professor Glantz asked the workshop participants to reflect on the evolving expectations of privacy, and whether there is still such a thing as a “reasonable expectation of privacy,” given our new data access and sharing landscape.

Dr. Rivera and others pointed out that conceptions of privacy change over time with new technology, and that our current system of research protections does not acknowledge the way people increasingly live their lives online and on social media where they are willing to accept reduced privacy for better online interactions.

Ms. Odwazny argued that the Common Rule’s regulatory scheme does not encourage this kind of research exceptionalism because “the regulatory definition of minimal risk does allow for daily life risk standards.” Furthermore, IRBs are often presented with big data research proposals that ought not to be regulated as human subjects research because the data being used is not private and identifiable. One possible solution, therefore, is IRB education of the type that PRIM&R provides, to help IRBs better understand when big data research is human subjects research, and when it poses more than minimal risks.

Dr. Ajunwa took issue with the idea of people “willingly trading their privacy,” which had come up earlier, because it suggests deeper awareness of how privacy is being invaded than many people have. For example, some employers tout their workplace wellness programs as benefits to employees, but don’t acknowledge that the programs are means to acquire employee’s health data for un-disclosed third party commercial uses.

Dr. Zimmer raised the point that social media giants like Facebook socialize people to share more information online and revise their privacy expectations. However, just because the public is increasingly more likely to share their information via social media platforms does not necessarily mean their expectations of privacy have changed; we should look at the context within which people originally chose to share their information.

Part III: Implications for consent, authorization, and data stewardship

The moderator, Jeremy Sugarman, MD, MPH, MA, of the Berman Institute of Bioethics at the Johns Hopkins University, introduced Part III by asking the panelists to address informed consent and potential alternatives to it in the new data access and sharing landscape.

Alex Capron, LLB, of the University of Southern California, examined some historical understandings of informed consent, including the conceptualization of consent as a means to protect subjects from harm, which has been in play in much of what we’ve already heard. The Belmont Report described informed consent as a way to serve the value “respect for persons,” where “persons” are, at least in part, moral agents capable of making autonomous choices

about their own lives and what they believe will lead to flourishing. The value of respect for persons thus goes beyond notions of privacy and physical integrity; an interference with someone's ability to make a choice may not be a harm, but is an affront to someone's dignity and therefore a wrong. He suggested that a reason informed consent continues to be an issue is that researchers, physicians, and others have for 40 years seen it as a one-time event thing, rather than an ongoing process that depends on mutual decision-making—a notion that has been codified in the consent requirements in the Common Rule. We should be thinking about mechanisms for securing the buy-in and authorization of the many people whose data are going to be used, accompanied by a mechanism that would give them assurance that the process is protective of their interests and of their contributions.

[Laura Odwazny](#), MA, JD, of the US Department of Health and Human Services, examined and summarized the mixed decisions and mixed messages US courts have issued on the question of whether informed consent documents should be considered contracts that would create legally binding obligations on the part of the parties to whom the consent applies (namely, the sponsor, the institution, the researcher, and the research subject). She further noted that the revised Common Rule includes provisions potentially relevant to data sharing that will be required to be included in both study-specific and broad consent. As the Office for Human Research Protections has the authority to enforce the promises made in the informed consent process that relate to the regulatory requirements, they can enforce these promises, specifically. Thus, if the final research data to be shared constitutes identifiable private information, the promises made to subjects in the informed consent process may limit the future uses of the data. For example, broad consent must address whether the researcher is planning to share identifiable private information collected from subjects, and for what general research purposes the information might be used; and because the newly added exemption for secondary research use conditioned on broad consent requires that the IRB determine that the research to be conducted is within the scope of the consent, the IRB will need to examine whether the contemplated research uses match the promises made to the subject.

[Rebecca Li](#), PhD, of the Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard University (MRCT Center), shared MRCT's findings from a small study looking at 12-15 existing consent forms and what they communicate about data sharing. About half of the forms were silent on whether data may be shared for future research purposes; some stated that data will be shared for very narrow purposes; and in a few instances, forms stated that data would be broadly shared with other researchers, although this broad language was at times conflicting. As a result, MRCT's multi-stakeholder working group—which included IRBs and patients and patient advocates—came up with recommendations about how to prospectively address data sharing in consent forms. They proposed definitions of terms such as “coded data,” “personal data,” and “de-identified data,” and created an informed consent template.

John Wilbanks, Chief Commons Officer of Sage Bionetworks, discussed his organization's efforts to re-design the informed consent process for shared data research using mobile apps. He observed that informed consent documents have been forced to take on roles they were not intended to play, and that people do not understand what is happening in the consent encounter. He and his colleagues approached consent as a “user-centered design project” (rather than an ethical or legal project) that anticipated not just consent, but also data sharing and returning information back to the participants. He argued that as we re-design consent for the future, we need to take seriously that consent is quite literally “the front door of the study.” He used examples to show that, if you take design as an opportunity and treat the mobile phone as a way to talk to potential participants, you can use design to make the study transparent and to help you in recruitment, engagement, and retention. With respect to data sharing, the app they designed asks people to choose between two options: sharing their data with just the

current researchers, or sharing it also with qualified researchers worldwide in sort of a controlled trust. Seventy-five percent of the people choose to share their data with researchers worldwide.

[Neal Dickert](#), MD, PhD, from Emory University School of Medicine, suggested that there is a good case for arguing that the public has a limited obligation to participate in research because medical knowledge is a public good and research is crucial to quality health care, which we all need. We should consider making sharing one's data for additional research purposes obligatory, especially in the context of record-based research data, which poses few new privacy and security risks. To be sure, studies show that people generally want to give their permission to have their data used for research; we shouldn't dismiss the association of informed consent with values such as trust and integrity. However, studies also show that people's preferences seem to be sensitive to practical limitations to getting consent. All of this suggests that we have an obligation to explore alternatives to the traditional informed consent model such as opt-in strategies and positive framing. Different models can predictably affect the decisions people make while still giving them control. There are many ways to advance transparency and trust, and we need more empirical scholarship to find the best approaches.

Dr. Sugarman then asked the panelists about how we should address the minority of participants who don't want their data shared, and whether that should affect the proposal that we re-evaluate the traditional informed consent model in which an individual is able to definitively make a decision about his or her participation. The following themes emerged in the course of discussion: the risks and harms that vulnerable populations face, improving partnerships and communication between researchers and participants, and the changing role of informed consent including new alternatives.

Dr. Li suggested that in clinical research, patients should be informed of data sharing practices, but that for the most part permitting secondary research by the same set of researchers that is collecting research data should be a condition of primary study participation, with some exceptions, for instance, what happened with respect to the Havasupai Indian Tribe's biospecimens. Concerns about the minority of patients who do not want to participate in research should be focused on populations that face greater risks such as the rare disease population whose data are more likely to be re-identified. Mr. Wilbanks pointed out that his organization changes their information modules depending on the type of data being collected, so that more sensitive data such as a person's genetic sequence will receive greater protection.

Dr. Dickert pointed out that data show not that people don't want to participate in shared data research, but rather that they prefer to be asked before their data are shared. Research participants might feel more like partners with researchers if communications around data sharing did a better job of fostering trust and community and gave people more "off ramps" if they do not want to participate. Professor Capron suggested that people want to be asked not because of their concerns about potential harms, but rather because of dignitary concerns and their desire for researchers to acknowledge the participant's role in the research process.

Ms. Okun said we should examine the purpose, the intent, and design of the research project to determine what kind of informed consent is needed. She cautioned that informed consent should not simply be disregarded in this new environment, but rather that the research oversight community should work on creating better partnerships between researchers and research participants. The patients her organization works with want to fully understand what they are being asked to participate in; they want to know not only the negative aspects of study participation but also the positive aspects, and they want the opportunity to ask and receive good answers about the nature of the research, recognizing that all future uses of their data might be unknown at the time of the study.

Dr. Dickert added that as we re-evaluate the research consent process, we should consider how we all too often fail to emphasize the positives of data sharing. Many potential participants may actually be supportive of such initiatives if they truly understood not only the negatives but also the positive aspects of research with shared data.

Stephen Rosenfeld, MD, MB, of Quorum Review IRB, noted that much health care delivery research is still commercially driven, and by not acknowledging that fact, we may lose public trust. Dr. Rivera noted that societies with single payer health care systems are in a better position to argue that participants' data should be shared when they receive care. Dr. Rubel pointed out that while people may be willing to share information as long as they are asked, we should also consider data collection situations that people have technically consented to but may be ethically problematic overall.

Mr. Wilbanks also pointed out that, based on his experience working with Apple on their ResearchKit, it seems the overwhelmingly majority of people who are already enrolled in clinical research are inherently altruistic and willing to share their data if they are asked what their preferences are on the matter. It is a mistake to see informed consent as something that can live solely in a Word document; his design features encourage people to see consent as more of a "long term conversation" between the researcher and participant, and allows participants to see how their data are being used and to withdraw from the research if they don't like the endpoint.

Dr. Rivera pointed out that while asking people about basic consent issues may be easily incorporated into emerging online and social media research designs, what is more tricky is how researchers follow through on promises that were made to participants about how their data will be used, especially when research participants are given the options to tier their choices. She thinks researchers should not over-promise to participants about the ways in which they can control how the data will be used.

Ms. Odwazny pointed out that consent is not only a way to improve transparency around data sharing; it also forms an obligation on behalf of the researcher to the subject, and some may think this obligation should be enforced either through legal mechanisms or through regulatory liability. However, we have not established what the first research institution's responsibilities are when there is downstream use of the data by a third party that is inconsistent with what was in the informed consent document.

P. Pearl O'Rourke, MD, of Partners HealthCare Systems, suggested that there is no common approach to the different kinds of data sharing initiatives, which may in the long run confuse the very participants researchers are trying to recruit, especially if they are enrolled in more than one study and each study has a different data sharing plan. We thus need to come up with a transparency plan that will work for multiple types of studies.

Dr. Fisher expressed concern about a positive exceptionalism approach that would justify waiving consent requirements across the board, and argued we should not abandon informed consent for any downstream use of data without considering what that means for public awareness of how the research establishment operates. She suggested any plan to waive informed consent in such situations be considered alongside a new, accountable "system of public transparency."

Dr. Hurley pointed out that much of the conversation seems to be about consent versus no consent, when we should also consider whether there are alternatives to traditional informed

consent model that can fulfill some of the functions we think consent should fulfill, or that might allow us to honor the values that are associated with informed consent, such as participant engagement and education, or transparency. Dr. Sugarman noted that in his empirical work around pragmatic clinical trials they are using “notification and authorization” terminology to capture a broader range of options than traditional consent for informing subjects and honoring preferences.

Dr. Strauss agreed with the other workshop participants that we need to consider a large cultural shift away from protectionism towards partnership, but cautioned that the revised Common Rule does not necessarily reflect this new engagement approach, because if a participant refuses to give broad consent for the future use of his or her identifiable biospecimens or data, researchers may de-identify the data or biospecimen and proceed to use it for research. Ms. Odwazny pointed out that this change mainly affects participants who may be dissuaded from participating in the primary study because they are not comfortable with the possibility that their data would be shared or used for research they might find objectionable.

Mr. Wilbanks said his organization engages and assesses research participant’s comfort level with future unknown uses of their data by giving them sample situations to see what kinds of benefits and risks they are comfortable with, and letting them know the results.

Dr. Rivera wondered if instead of coming up with a whole new system of rules to facilitate improved data access, use, and sharing, it might be better to make adjustments to the system we currently have by making it more communitarian, so that patients become accustomed to donating their data or biospecimens without question in a clinical care setting. Professor Capron argued that we also need to consider the logistical difficulties that were raised in Part I, including the fact that researchers and clinicians are not yet ready to fully embrace a system in which it is assumed that data will be shared.

Part IV: Responsibilities for ethical oversight of research involving data access, use, and sharing

Albert J. “A.J.” Allen, MD, PhD, of Eli Lilly and Company, introduced Part IV by noting that within the context of the ethical oversight of research involving data access, use, and sharing, we are particularly interested in the roles and responsibilities of different groups.

[John R. Baumann](#), PhD, of Indiana University, pointed out that research data are typically owned by the institution, not the researcher or the study participant. When researchers share their data with people external to their institutions, or researchers leave, institutions struggle with questions about who maintains responsibility for the data. While the IRB has the central role in the regulatory and ethical review of data sharing, it cannot be the only entity exercising responsibility over data. Institutions need an integrated approach that involves understanding and taking responsibility for the data they own, and identifying other components of their institutional data oversight community. Institutions should also develop policies around the access, use, and sharing of their datasets that address the different types of data including whether it is being shared for commercial or non-commercial purposes, how sensitive it is, and whether it is identified or, if not, can be easily re-identified. Although we should trust researchers to keep their promises about what they are going to do with data, institutions should consider using not-for-cause monitoring, and should assertively respond to allegations of inappropriate sharing of data.

[Brian Herman](#), MD, of the University of Minnesota, spoke about the institutional conflicts of interest (ICoIs) in human participant research, which are becoming more prevalent as academic institutions increasingly follow big business models and see their data sources as their intellectual property. Institutions would like to commercialize that data themselves as a revenue source, rather than sell the data to a third party or give it away for free. These issues historically are managed by institutional conflict of interest committees, which are part of a self-reporting system permitted by the federal government. But Dr. Herman is beginning to wonder how independent these bodies are from the institutions they are supposed to watch over. We should look into “extra-institutional approaches” to managing ICoIs to regain public trust in institutional research and encourage people to be more willing to share their data in that context, and he suggested we look, specifically, to professional societies, the state or federal government, or independent organizations such as PRIM&R or AAHRPP to help.

Deven McGraw, MPH, JD, of the Office for Civil Rights, US Department of Health and Human Services (HHS), speaking on behalf of herself and not the Administration, began by pointing out there are many entities collecting valuable data who are not yet sold on the value proposition of making that data available for research purposes. Entities covered under the Health Insurance Portability and Accountability Act (HIPAA) are permitted but not required to disclose data for research purposes, unless an individual exerts his or her right to have a copy of his or her health information transmitted to a third party under the “right of access” provision; interestingly, this is the provision on which the national “All of Us” shared-data research program relies. The 21st Century Cures Act mandates that HHS issue more guidance in this area and review what constitutes “information blocking,” and whether it includes not making data available for legitimate research purposes. She also pointed out HIPAA promotes research exceptionalism, in that it allows research data to be shared for quality improvement purposes, but requires data sharing that is to contribute to generalizable knowledge to follow a more stringent set of rules. In her own personal opinion, it might be worth re-considering whether data should be regulated on the basis of privacy risk rather than based on the purpose for which the data are being used.

Speaking on behalf of the British Medical Journal, a member of the International Committee of Medical Journal Editors (ICMJE) Elizabeth Loder, MD, MPH, built on her earlier comments (in Part I) by saying she is not sure that journals are in the best position to adjudicate the variety of issues that arise in the research process, beyond publication, because they are not always subject matter experts and often are dependent on information supplied by researchers and others—for instance, assertions that IRB approval has been obtained. However, she supports the idea that journals can provide an extra-institutional check on data sharing and research practices, especially because institutions themselves are often conflicted or are not very good at investigating themselves. Journals can adjudicate issues by publishing or refusing to publish, issuing corrections and notices of concern, contacting funders, and causing embarrassment and inconvenience. They can also reserve the right to “audit” ethical aspects of the study, such as IRB approval and informed consent, though they currently don’t do that. However, while larger journals might be interested in assuming an impartial oversight role, smaller subspecialty journals, which publish most research, are not in a position to do this.

Dr. Allen then asked the panelists to comment on additional groups, such as investigators, that should play a role in fostering responsible data access, sharing, and secondary use, and what kind of guidance is needed for them.

The following themes emerged in the course of discussion: The need for better institutional support, resources, and education for both researchers and oversight bodies; the importance of “good principles of data stewardship” and transparency; the need to re-evaluate how to gain participants’ trust; and the role of the IRB and other oversight entities in this new landscape.

Dr. Herman made the point that investigators often need institutional support to create and maintain shared research databases in a secure manner, which is why at the University of Minnesota they created an Office of Research Computing to assist them, and in the process learned that the Office could also help faculty use pre-existing datasets for new research purposes.

Dr. Baumann agreed, as an institutional official, that researchers need additional tools and guidelines, but cautioned that institutions are loath to introduce new requirements unless there is a clear regulatory basis for them. He also believes that most researchers do not know how to re-identify data, but also don't understand that it is unethical to do so, and shouldn't be attempted. Dr. O'Rourke mentioned that many in the research sphere do not understand what de-identification and related terms mean.

Ms. McGraw argued that researchers also need "good principles of data stewardship," and that the informed consent process cannot be the only vehicle for making sure that data are being used responsibly and in an ethical manner. Researchers must be taught that they should not collect more data than needed to meet their goals, and that data should not be retained indefinitely. Although we may not be able to completely eliminate the threat of re-identification, there are security measures we can take to more successfully manage it. Dr. Allen mentioned that several European researchers have already begun work on a code of conduct for researchers working with secondary datasets and the data stewardship issues they face.

Dr. Loder agreed with Dr. Allen about the importance of investigators working with journals to raise the bar of research standards, and pointed out that during the peer review process it is often investigators who identify ethical lapses such as improper use of datasets. She believes journals might also be able to showcase best practices and the positive outcomes of shared data research, in addition to publicizing ethical lapses.

Workshop participants then discussed how researchers and ethical oversight bodies can facilitate shared data research that participants can trust.

Dr. Baumann gave an example of a Kenyan bio-banking project in which local physician researchers were at first reluctant to share specimens because they assumed they would be sent to American researchers; however, the institution worked to build their trust with an educational campaign that highlighted new governance rules and how this effort would be different from past bio-banking projects in Africa. Dr. Herman agreed that being open about the process, including the benefits of shared data research for individuals, is crucial, but at the same time such efforts to build trust may be for naught if institutions do not hold people accountable for ethical violations.

Dr. Rosenfeld noted that trust often depends on context, and we need to remember that people often agree to be in clinical trials because of their doctor's recommendation; thus, while the institution may play a larger regulatory role in facilitating the sharing of data, we should be mindful of this personal relationship.

Ms. McGraw said that although the largest data breach in the US did not occur in a research environment, researchers' should strive for a shared data research environment that participants can generally trust, and in which they are willing to share their data for as long as it is valuable, even if the enterprise is not 100% risk free. She highlighted that earlier in the day, panelists described innovative ways to improve transparency, including researchers listening to and working with communities before designing research projects so that participants

understand the benefits and risks and feel part of the overall process. In her personal opinion, she urged that we go beyond "check-the-box exercises" that don't leave a lot of "mental room for innovative design" of projects that might work better from the standpoint of transparency and trust, and that we then evaluate how well these projects engaged participants in the process.

Workshop participants also discussed what kind of governance is needed to oversee research with shared data and promote transparency amongst stakeholders who may have inherent conflicts.

Dr. Hernandez pointed out that even as the number of shared data research enterprises increases, including those with industry partners, there is still no oversight body that can establish a system of checks and balances and require relevant actors to appropriately disclose their processes, funders, and results.

Ms. Selwitz argued that the IRB community does not have the resources to deal with oversight and potential enforcement issues related to data sharing, especially if the community is not given adequate funding to handle additional workload issues. She asked the workshop participants to consider alternatives to the IRB (e.g. data safety monitoring boards and privacy officers), especially since in the future many institutions will be outsourcing review of the research at their institution to a single IRB. Dr. Baumann agreed that a separate privacy board might be able to provide oversight on data sharing issues, but cautioned that often other institutional authorities beyond the IRB do not want to assume additional responsibilities.

Ms. Hansen, of the Fred Hutchinson Cancer Research Center, agreed that the IRB is all too often the "institutional backstop" for a number of different study ancillary review responsibilities; for example, the federal government currently requires IRBs to evaluate genomic data sharing plans. She suggested that any messaging from this workshop should be brought to the attention of the highest institutional officials. She also proposed that a third-party accreditation process be set up to review how institutions share data and establish best practices.

Dr. Herman said the University of Minnesota created a separate committee composed of computer and biomedical scientists to review data management and sharing plans for the IRB; that committee ultimately makes the decision about whether or not to approve the research; this reduces IRB burden and allows researchers with expertise to weigh in.

Professor Glantz argued that one reason IRBs might be tasked with data sharing responsibilities is because authorities still mistakenly assume that research with shared data should be treated the same as research done directly on humans.

Dr. Herman floated the idea that perhaps an independent nonprofit could be set up to objectively address these conflicts, set policies, and take feedback from relevant stakeholders in a manner that the public could trust since the nonprofit would not benefit from the outcomes of the policies they set. Currently, institutions set their own policies, which is problematic as they are also under pressure to generate income in a competitive research environment.

Conclusion: Summary and next steps

Professor Capron reviewed the day's discussion by noting that the barriers to using data that are collected for one purpose for other research purposes go beyond practical and logistical concerns. Such problems include the lack of incentives for various stakeholders to share or analyze pre-existing data, and institutional cultures that do not encourage the re-use of data

despite the advantages for public health and patients. He noted that we learned of the potential promising uses of data, including health services data, about how much patients who are informed about data sharing embrace and welcome it, and about the need to address the issue of reproducibility in science.

He concluded that it is clear that we need to move in the direction of enhanced data sharing. Throughout the day, we heard about innovative ways to resolve a “conflict” between the interests of individuals and those of the community. But it seems that if we can figure out a mechanism for doing so, we should be able to avoid pitting individual and community interests against each other (i.e., such that either individuals will be harmed or their privacy invaded, or we are going to lose out on these community goods), and arrive at a reasonable accommodation that does not ask individuals to step out of the picture and be uninvolved sources of data and nothing more.

A key ingredient we have heard about today is trust, specifically, “justified trust,” trust that rests upon a system that not only entails security of data, but is also aimed at the common good, where we understand a truly “common” good to be both just and fair.

What mechanisms make sense for establishing a form of responsible stewardship that the public as a group as well as individuals asked to share their data can rely on? Picking up on Ms. McGraw’s suggestion that we need “mental room for innovative design” and Dr. Dickert’s suggestion that there may be a limited obligation to participate in research, Professor Capron suggested that we consider in the data sharing space an analogy to what happens in public health.

Within the public health framework, physicians have responsibilities both to the care of their individual patients, and to the good of the community. If I have suspicious symptoms and the doctor takes a sample and sends it to the lab even though I do not want her to, it is not because she is overriding my choices about my treatment; it is because she has obligations to the community as well. And when individuals have complaints about how they are treated, or how the system is working, there is public debate. Public health surveillance is an activity that is identified as a public responsibility and a public process under the civil authorities. If we moved toward this model with respect to data sharing, the kinds of conversations that we had today about whether the prize that we’re seeking is likely to be found and whether it is worth some cost in reducing our otherwise high expectations for having individual choice, could happen in a public way, with the participation of various stakeholders.

Dr. Rosenfeld lamented that we seem to do everything by taking the system we have and throwing more and more requirements on it in a reactive fashion every time something new comes up in research. So he is excited about the opportunity to remove some of these barriers if we pursue something like Professor Capron’s suggestion, that is, if we take all of the requirements and reengineer them in a research care delivery system from the ground up. Perhaps a demonstration project is the place to start.

Dr. Rodriguez asked workshop participants to encourage the research oversight community to start a dialogue on differential risk and what’s acceptable for different kinds of research with shared data, especially when you consider things like cultural and logistical factors.

Ms. Selwitz argued that in order for there to be a real culture shift, investigators need more education on why research with shared data is important.

Dr. O'Rourke argued that many of the analyzed data sets we might be interested in sharing are of poor quality and that researchers erroneously consider things like electronic medical records valuable data sources. Dr. Rivera pointed out that maybe we need to first figure out what data sources are worth sharing, and ensure, in light of scarce resources, that extremely valuable data are being shared. Dr. Ross argued that large payers and funders of research (in both the public and private sectors) are increasingly recognizing the value of research with shared clinical data and coming up with infrastructures to maximize data sets that are deemed to have the most value, in part because this allows them to address reproducibility issues. Dr. Rosenfeld pointed out that standards and designs for sharing data should be developed with the researcher in mind to encourage them to do a better job of using pre-existing datasets.

Dr. Loder raised the point that meta-analysis is a valuable way to review data, but you still have to identify and evaluate the quality of the studies to be included, a problem for aggregating data sets in general. Dr. Strauss thought we might be able to address this in part by studying data collection and sharing methods to identify problem areas, so that we can reduce the number of unused datasets.

Dr. Allen suggested that the research ethics community might come up with a set of common principles for research with shared data reflecting the concerns raised throughout the day including: community engagement; making research relevant to communities; improved education of key stakeholders; some way of honoring people's desire to be asked regarding collection and use of their data; and clearer, more consistent definitions.

Dr. Strauss added that the day's proceedings showed the enormous opportunities that research with shared data pose for advancing science and health care, as well as great need and value of engaging various stakeholders about the best way forward. The stakes are high, and the path to get where we need to be not a simple or straightforward one. We learned we don't have systems, models, or regulations in place that help us with the demands we are facing around data stewardship, around investigator education and training, around data management, and possibly around consent. In fact, we may need a fairly elaborate deconstruction of our way of thinking about some of these things. Some of what we need to know we can learn through empirical work with investigators and with participants of research. And we will need to deal with barriers that have to do with the increasingly decentralized approach to managing some of these issues, as per the new Common Rule.

Dr. Hurley closed the day by thanking participants and letting them know that PRIM&R would be in touch about next steps and future work products.