

Data Sharing and Pragmatic Clinical Trials: Law & Ethics Amidst a Changing Policy Landscape

Kayte Spector-Bagdady JD MBe

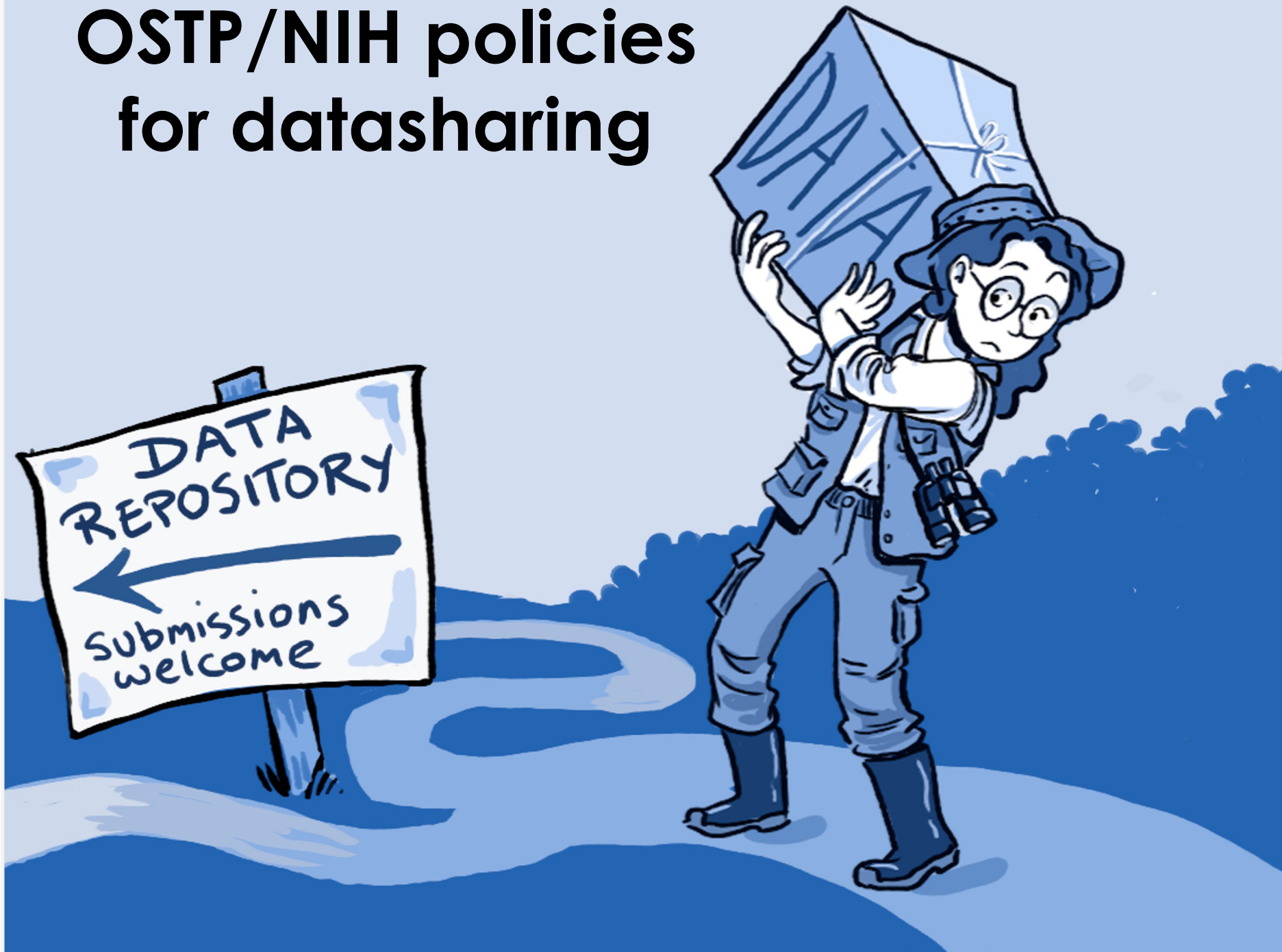
Associate Director, Center for Bioethics & Social Sciences in Medicine

Assistant Professor, Obstetrics and Gynecology

NIH Collaboratory Grand Rounds Ethics and Regulatory Series

November 11, 2022

OSTP/NIH policies for datasharing



2003 NIH Data Sharing Policy and Implementation Guidance

- **Scope:** PIs asking for \$500,000 or more in direct costs per year
- **Ask:** Plan for sharing final research data for research purposes, or state why data sharing is not possible
- **Data sharing:**
 - Timing: No later than the acceptance for publication of the main findings from the final dataset
 - Content: Must include information necessary to “document, support, and validate” findings as well as avoid “misuse, misinterpretation, and confusion”
- **Of note:**
 - Investigators have legitimate interest in benefitting from time and effort – first use but not prolonged exclusive use
 - Concerns re co-funding with industry, wanted to protect proprietary data

2013 OSTP Increasing Access to the Results of Federally Funded Scientific Research Memorandum

- **Scope:** 20 federal departments and agencies with over \$100 million in annual R&D expenditures must develop a plan
- **Ask:** 12-month embargo period for papers “directly arising from federal funding” (whole or in part)
- **Of note:**
 - **Publishers** provide valuable services, including the coordination of peer review, that are essential for ensuring the high quality and integrity of many scholarly publications. It is critical that these services continue to be made available.
 - It is also important that Federal policy not adversely affect opportunities for **researchers who are not funded** by the Federal Government to disseminate any analysis or results of their research.
 - Want to protect private interests and encourage cooperation via **public/private partnerships**

2014 NIH Genomic Data Sharing Policy

- **Scope:** focused on genomic data “generated” from NIH funding
- **Ask:** PIs must share genomic data, including the data necessary to interpret, in an NIH-designated repository by the time of publication of their first article using the data
- **Consent:** requires investigators to request informed consent for future use and sharing of genomic data derived from (even de-identified) cell lines or clinical specimens collected after effective date
 - “it is increasingly clear that participants expect to be asked for their permission to use and share their de-identified specimens for research,” even if those specimens are de-identified as defined by the HIPAA Privacy Rule (e.g., lacking name or address).

2020 NIH Policy for Data Management and Sharing

- **Scope:** From \$500,000 in costs to *all research, funded or conducted in whole or in part by NIH, that generates scientific data*
- **Ask:** Maximize data sharing through informed consent process
- **Data sharing:**
 - Timing: No later than the acceptance for publication of the main findings from the final dataset or *end of award* – *whichever is 1st*
 - Content: Data necessary to both validate and *replicate* findings, even if not published
- **Of note:**
 - Will update the 2014 GDS policy as well
 - Effective January 25, 2023

2022 OSTP Ensuring Free, Immediate, and Equitable Access to Federally Funded Research Memorandum

- **Scope:** D&A with over \$100 million in annual R&D to those with *any*
- **Ask:** all published articles resulting from federal funding (including funding held by co-authors) be made “freely available and publicly accessible” without embargo or delay
- **Data Repositories:** should provide free and easy access, curation and quality assurance, common formatting, clear provenance, and fidelity to consent
- **Of note:**
 - **Equity:** Responds to years of public feedback that 12-mo embargo was inequitable
 - SOS develop measures to additionally reduce inequities for “individuals from underserved backgrounds and those who are early in their careers,” as well as reduce the burden of data sharing on funded researchers generally
 - **Transparency:** Surrounding the generation of federally funded scholarship, including “authorship, funding, affiliations, and development status” of the work.

Three thoughts...

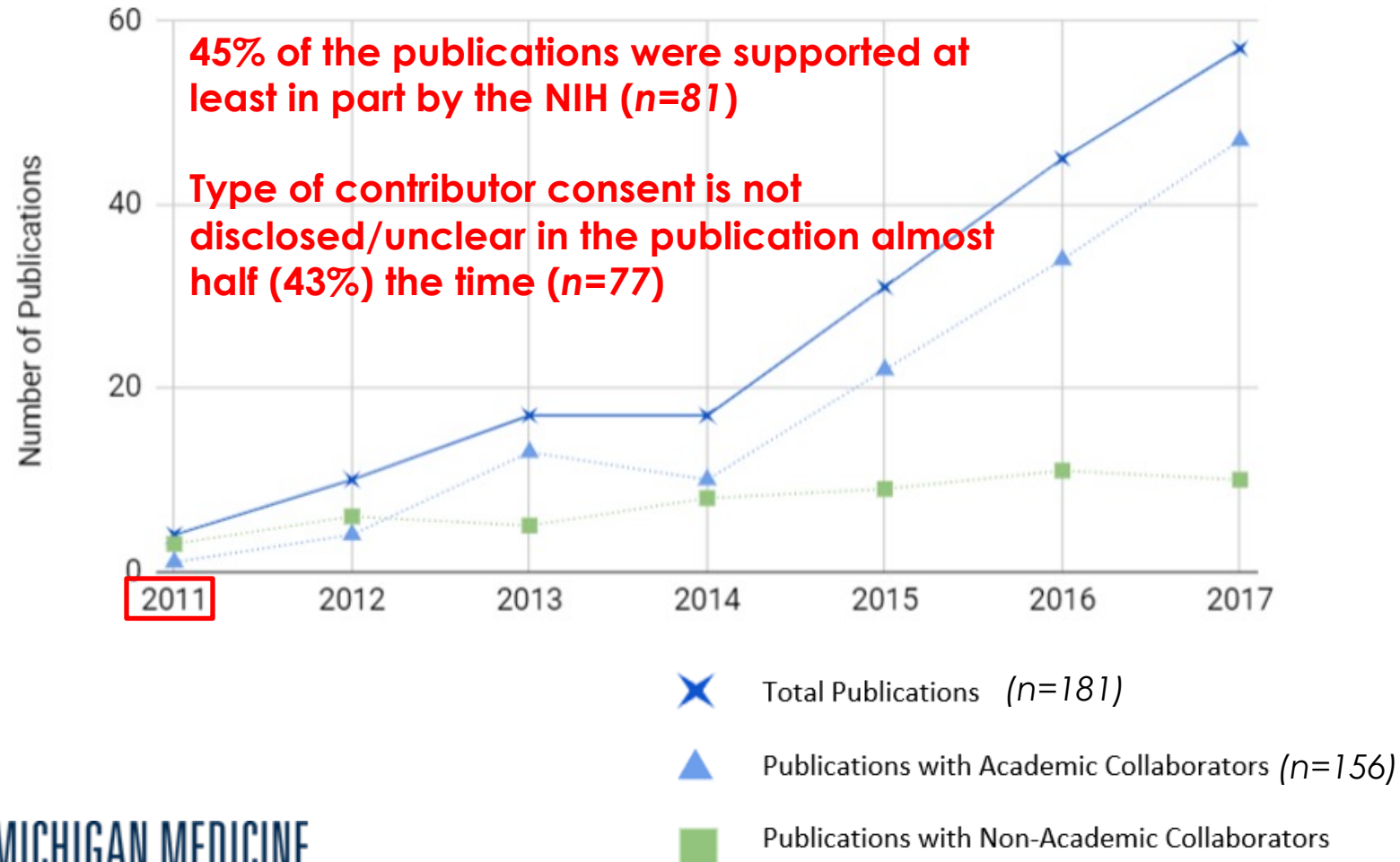
Sharing data was seen as a burden without reward

*“Keeping our labs motivated, keeping our post docs motivated, **keeping them productive is hard enough** and then having [to make] them go through some **really cumbersome process** to make their data available, which involves both bureaucratic work and work organizing and curating the data, which **people don't often see benefit from?** So, yeah, I think it's a lot of things that make [data sharing] challenging.”*



Genetic data partnerships: academic publications with privately owned or generated genetic data

Kayte Spector-Bagdady, JD, MBE¹, Amanda Fakhri, MHSA², Chris Krenz, BA³,
Erica E. Marsh, MD⁴ and J. Scott Roberts, PhD⁵



Public private partnerships

- Relationship governed by contract – PIs cannot share back to government datasets
- Publishing with industry data adds value to that business asset
- Current GDP limited to funding used to “generate genetic data” but preliminary survey results show that the #1 thing industry data researchers use federal funding for is to *analyze data*

“...it **legitimizes [23andMe] as a company, makes them look better in their research**....they get their genetic insights followed up on and they prove that **the way they collect data is valuable**, mainly by the self-report, and maybe that helps them build a case for then **selling the data** to various drug development companies.”

“...this became a really important **roadblock for us in terms of publishing** the paper, because basically, the journal said, ‘Your paper's interesting. We would love to see a revision. But you need to make the data available.’ And 23andMe said, ‘**Well, we can't do that.**’”

are encouraged to attend. NIH has established a 45-day public comment period for the scoping process.

Dated: August 21, 2014.

Daniel G. Wheeland,

*Director, Office of Research Facilities
Development and Operations, National
Institutes of Health.*

[FR Doc. 2014-20489 Filed 8-27-14; 8:45 am]

BILLING CODE 4140-01-P

DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Final NIH Genomic Data Sharing Policy

SUMMARY: The National Institutes of Health (NIH) announces the final Genomic Data Sharing (GDS) Policy that promotes sharing for research purposes

The NIH received a total of 107 public comments on the draft GDS Policy. Comments were submitted by individuals, organizations, and entities affiliated with academic institutions, professional and scientific societies, disease and patient advocacy groups, research organizations, industry and commercial organizations, tribal organizations, state public health agencies, and private clinical practices. The public comments have been posted on the NIH GDS Web site.¹⁰ Comments were supportive of the principles of sharing data to advance research. However, there were a number of

Allergy and Infectious Diseases, Sharing and Release of Research Data program requirements, definitions of key terms, and the Policy (e.g., aggregated data included and other clarifications).

The statement of intent to intentionally generate genomic data to accommodate the e range of research the data. It also allows that individual NIH (IC) may choose on to apply the Policy

identification.¹⁹ Moreover, requiring that consent be obtained is respectful of research participants, and it is increasingly clear that participants expect to be asked for their permission to use and share their de-identified specimens for research.^{20, 21, 22} The

- **D.J. Kaufman et al. ‘Public Opinion about the Importance of Privacy in Biobank Research,’ (2009): 643-654:** $\frac{3}{4}$ were concerned about “the government having [their] samples and information,” and 56% were concerned about “researchers having [their] samples and information.”
- **E. Vermeulen et al. ‘A Trial of Consent Procedures for Future Research with Clinically Derived Biological Samples’ (2009):** the majority of respondents (72%) expected to be informed about research findings based on the use of their tissue
- **S.B. Trinidad et al. ‘Research Practice and Participant Preferences: The Growing Gulf.’ (2011):** the Common Rule doesn’t cover de-identified specimens or data

Table 2. Participant Preferences for Notification About Use of Identified or Deidentified Health Information and Biospecimens by Commercial Companies and University Researchers (N = 2054)

Question	For you, how true are the following statements: I would like to be notified about [I] using my [II] [III].			Mean (95% CI) ^a	P value
	I	II	III		
A	Commercial companies	Identified	Biospecimens	3.45 (3.41-3.50)	
B	Commercial companies	Deidentified	Biospecimens	2.94 (2.89-2.30)	
C	Commercial companies	Identified	Health Information	3.45 (3.41-3.49)	
D	Commercial companies	Deidentified	Health Information	2.96 (2.91-3.01)	
E	University researchers	Identified	Biospecimens	3.37 (3.33-3.41)	
F	University researchers	Deidentified	Biospecimens	2.76 (2.71-2.81)	
G	University researchers	Identified	Health Information	3.44 (3.40-3.48)	
H	University researchers	Deidentified	Health Information	2.77 (2.72-2.83)	
Comparison of composite Indexes of preference for notification^b					
Index 1		Index 2		Mean difference (95% CI)	
Questions used in Index	Mean (95% CI) ^a	Questions used in Index	Mean (95% CI) ^a		
All commercial companies (A, B, C, D)	3.20 (3.16-3.24)	All university researchers (E, F, G, H)	3.09 (3.05-3.12)	0.12 (0.09-0.14)	<.001
All identified materials (A, C, E, G)	3.43 (3.39-3.46)	All de-identified materials (B, D, F, H)	2.86 (2.80-2.90)	0.57 (0.52-0.61)	<.001
All health information (C, D, G, H)	3.15 (3.12-3.19)	All biospecimens (A, B, E, F)	3.13 (3.10-3.17)	0.02 (0.00-0.04)	.01

^a Mean responses on the following Likert scale: 1 = not true; 2 = somewhat true; 3 = fairly true; 4 = very true.

^b Indexes are generated as the sum of participants' answers to questions A through H divided by the number of questions answered.

Spector-Bagdady K., et al. Reported Interest in Notification Regarding Use of Health Information and Biospecimens. *JAMA* 2022;328(5):474-476.

thank you

Kayte Spector-Bagdady
kaytesb@med.umich.edu
@kaytesb 

Ethical Challenges for Sharing Data from Pragmatic Clinical Trials

Stephanie Morain, PhD, MPH

November 11, 2022



Why PCT Data Sharing “Different”



Often use waivers
or alterations of
informed consent



Embedded into ongoing
clinical care, using
extant data

Why PCT Data Sharing “Different”



Often use waivers
or alterations of
informed consent

Traditionally...

- data sharing presented as honoring the preferences of trial participants

Compiled Public Comments on a DRAFT NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance

Guide Notice Number: NOT-OD-20-013

November 06, 2019 – January 10, 2020

“It is especially worth considering that many human participants expect their data from their participation will be shared with other qualified researchers.”

“...clinical trial data sharing also respects trial participants’ assumption of personal risk to contribute to science by maximizing the value of their contributions.”

SPECIAL ARTICLE

Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing

Michelle M. Mello, J.D., Ph.D., Van Lieou, B.S.,
and Steven N. Goodman, M.D., Ph.D.

ABSTRACT

BACKGROUND

Sharing of participant-level clinical trial data has potential benefits, but concerns about potential harms to research participants have led some pharmaceutical sponsors and investigators to urge caution. Little is known about clinical trial participants' perceptions of the risks of data sharing.

METHODS

We conducted a structured survey of 771 current and recent participants from a diverse sample of clinical trials at three academic medical centers in the United States. Surveys were distributed by mail (350 completed surveys) and in clinic waiting rooms (421 completed surveys) (overall response rate, 79%).

RESULTS

Less than 8% of respondents felt that the potential negative consequences of data sharing outweighed the benefits. A total of 93% were very or somewhat likely to allow their own data to be shared with university scientists, and 82% were very or somewhat likely to share with scientists in for-profit companies. Willingness to share data did not vary appreciably with the purpose for which the data would be used, with the exception that fewer participants were willing to share their data for use in litigation. The respondents' greatest concerns were that data sharing might make others less willing to enroll in clinical trials (37% very or somewhat concerned), that data would be used for marketing purposes (34%), or that data could be stolen (30%). Less concern was expressed about discrimination (22%) and exploitation of data for profit (20%).

“most clinical trial
participants...believed that
the benefits of data sharing
outweighed the potential
negative aspects and were
willing to share their data”

Why PCT Data Sharing “Different”



Often use waivers
or alterations of
informed consent

Traditionally...

- data sharing presented as honoring the preferences of trial participants
- heavy emphasis on role of informed consent to fulfill the ethical obligation to respect those whose data are shared



Sharing Clinical Trial Data

MAXIMIZING BENEFITS, MINIMIZING RISK

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

“for most prospective trials...the informed consent process provides an opportunity to obtain participants’ approval for planned data sharing and to be transparent about potential future data sharing”

If PCT uses a waiver/alteration of consent...



- Cannot assume sharing data is consistent with preferences of patient-subjects
- Cannot rely on informed consent to fulfill ethical obligation of respect

What does it mean to respect patient-subjects in the context of (not) sharing data from a PCT conducted under a waiver/alteration of informed consent?

Gatekeepers as data stewards?



- IRBs/HRPPs
- Investigators
- Health system leaders

Yet divergent perceptions about patient-subject preferences...

*“...if people know that you’re doing research... for public good, and not for profit, **people are generally enthusiastic** about [their] participation being used by others to learn more.*

-Health System Leader

Yet divergent perceptions about patient-subject preferences...

“...if people know that you’re doing research... for public good, and not for profit, people are generally enthusiastic about [their] participation being used by others to learn more.

-Health System Leader

“...research participants want to be asked...when we talk about downstream sharing of deidentified data it would probably run along the same lines of whether or not people...would be bothered by the fact that they were in a study under a waiver in the first place....” –HRPP Director

...and are concerned about INSTITUTIONAL risks from patient-subject reidentification

“The nightmare scenario for some of the research centers is...some public embarrassing revelation about individuals being re-identified... [from data shared from a PCT using a waiver of informed consent]”

-PCT Investigator

Perceived need for greater public awareness of need for/benefits of data sharing...

“We’ve done a horrible job in this country educating people about the value of research, and [helping] people understand that putting constraints around our ability to use data as researchers also puts constraints around the change that we’re going to be able to find meaningful treatments to address people’s problems.”

--PCT Investigator



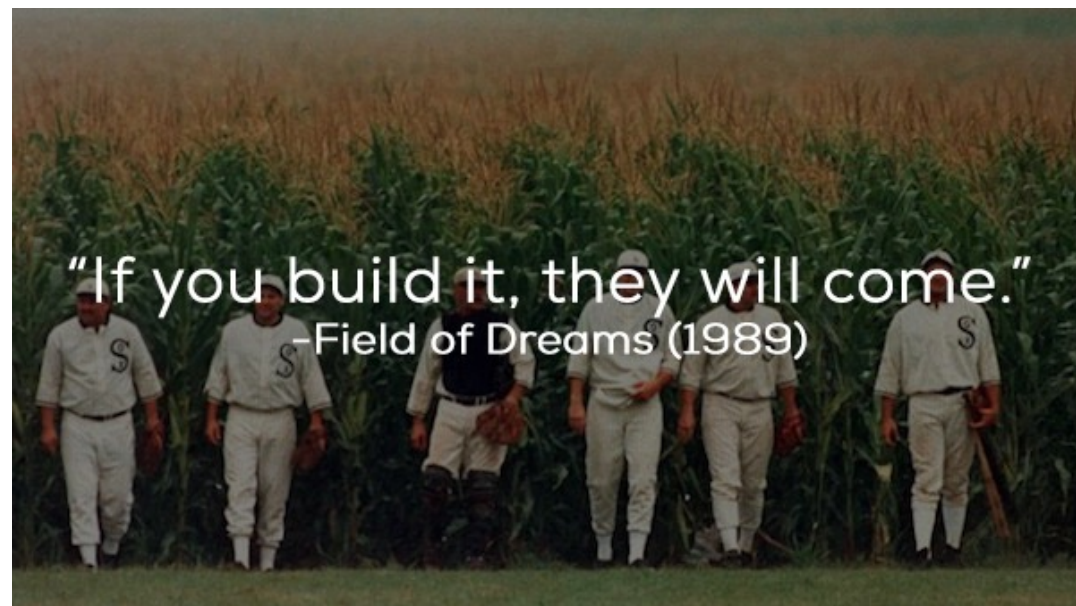
Sharing Clinical Trial Data

MAXIMIZING BENEFITS, MINIMIZING RISK

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

“...the public does not fully understand the benefits and value of data sharing, and the demand is not commensurate with the need for change.”

“Engendering support for data sharing will require greater awareness of how the use of electronic health care data has led to improved outcomes...”



"If you build it, they will come."
-Field of Dreams (1989)

Why PCT Data Sharing “Different”



Often use waivers
or alterations of
informed consent



Embedded into ongoing
clinical care, using
extant data

Implications of Embeddedness for PCT Data Sharing

- Data volume potentially larger (& substantially so)
- Data may be “about” those beyond patient-subjects
- Data may have been collective for administrative/clinical purposes
- Data may be more representative of “real world” conditions
- Data may be controlled by a third party (e.g, CMS)



Implications of Embeddedness for PCT Data Sharing

- Data volume potentially larger (& substantially so)
- Data may be “about” those beyond patient-subjects
- Data may have been collective for administrative/clinical purposes
- Data may be more representative of “real world” conditions
- Data may be controlled by a third party (e.g, CMS)



Benefits

Challenges

Larger Datasets → Greater Statistical Power

*“One of the potential benefits of the PCT is that they’re usually much **larger datasets**, which means there’s **lots of additional potential subanalyses**.”* –PCT Sponsor

RWD → Improved Relevance of Secondary Analyses

“...if I really want to know what’s going to happen in the real world ...then the PCT data is actually a lot more interesting [than efficacy data] ...even though it’s a much bigger burden to deal with...” --PCT Investigator

RWD Nature + Greater Volume of Data - > Increased Logistical Burdens to Prepare Data for Sharing



To ensure
DEIDENTIFICATION



To minimize risk of
BIASED/MISLEADING
ANALYSES

Increased Risks for Privacy Violations

- Greater potential for data to inadvertently contain identifiable info
- Enhanced logistical challenges to ensure deidentification
- Enhanced potential for reidentification through data linkage

Increased Risk of Biased/Misleading Analyses

*“When PCTs try to leverage EMR data, non-primary data, **an enormous number of issues show up around the provenance of that data.** What generated it? Why did it show up? How does it vary from site to site?*

*...the downstream use of that data, in my experience, is hard...**you need to have so much knowledge about the nuances of that data to ultimately use it well.**”* –Biostatistician

Disclosure about Systems/Clinicians

*“The other interest that impedes [data sharing] is **concerns of healthcare systems that they may be being compared one to another**, providers may be compared...”*

–Biostatistician

Proprietary Data/Health Care Operations

“Say, for example, you were aggregating the kinds of drugs that people were being prescribed as part of a clinical trial, and that got shared, and a competitor to a health system might be able to figure out that this is a prescribing practice, and maybe they got a great deal from the provider of that drug, they could use that to better their business interests. That could happen with drugs, devices—any element that’s bought and ends up in the medical record.”

—PCT Investigator

Data Owner Restrictions

“It's more difficult [to share PCT data] because this isn't data that you're collecting as a researcher in a lab and it's totally your data. In a pragmatic trial you're using existing health records, claims data, EMR data...the data's not yours to use, even from the beginning.”

--PCT Investigator

Is Sharing PCT Data “Worth It”?

Data Sharing — Is the Juice Worth the Squeeze?

Brian L. Strom, M.D., M.P.H., Marc E. Buyse, Sc.D., John Hughes, B.Sc., and Bartha M. Knoppers, Ph.D.

EDITORIAL

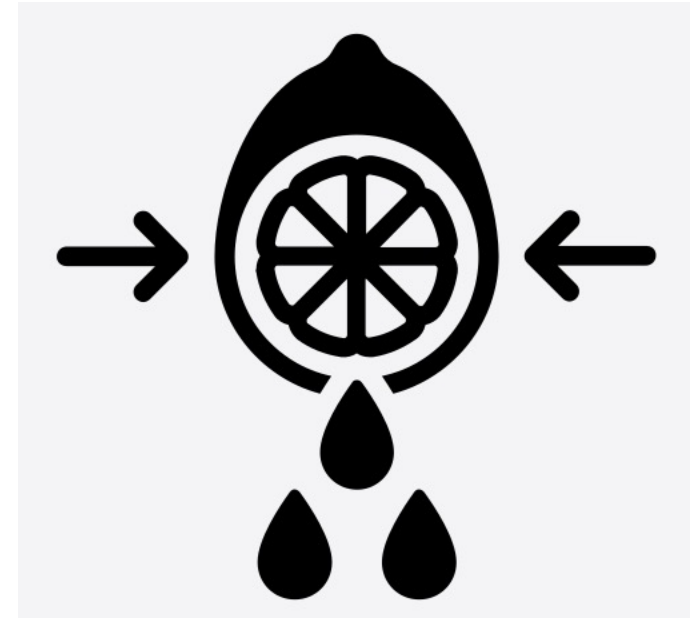
Data Sharing—The Time Has (Not Yet?) Come

Clyde W. Yancy, MD, MSc; Robert A. Harrington, MD; Robert O. Bonow, MD, MS

No Agreement on “Juice” or “Squeeze”



Benefits/Payoff?



Burden, Risks of Sharing?

Investigators/Health Systems: Substantial Risks & Burdens for Unclear Benefits

- Substantial logistical burdens in preparing data for sharing
- Meaningful risks of reidentification (of patients, health care systems)
- Concern for biased/misleading analyses
- Little demonstrated demand for PCT data
- Relatively low social value from PCT data reuse

“I think the amount of effort that goes into sharing...what comes out of it is hard to quantify and has a couple real wins but ...is not this massive source of value that people thought it might be... right now this is about methods innovation and about other things, but it's not really generating high-impact, important work”

—PCT Investigator

Sponsors

“The resources needed for repositories are actually quite modest compared to the resources needed to actually conduct the study ...you have an institute with a couple billion dollars and so it cost you \$3 million a year to maintain a repository of data...It's a very small percent of the total cost.”

—PCT Sponsor

“The investment to support data sharing is less than one percent...of the research spend, so just kind of on those grounds is kind of a no-brainer when you look at the long-term benefits... it's completely worth the investment.”

—PCT Sponsor

Suggested Takeaways?



Use of waivers or
alterations of
informed consent



Look beyond informed consent
processes to fulfill obligations
of respect when sharing
individual-level data from PCTs

Suggested Takeaways?

If public demand for greater data sharing is to be driven by awareness of its benefits...

Need to do a better job measuring—and communicating about—the “juice”





JOHNS HOPKINS

BERMAN INSTITUTE
of BIOETHICS