Ethics of sharing individual-level data from research with human subjects

Saskia Hendriks, MD, PhD
saskia.hendriks@nih.gov

Any views expressed are my own and do not represent those of the NIH, HHS, or the US Government.

No conflicts of interest

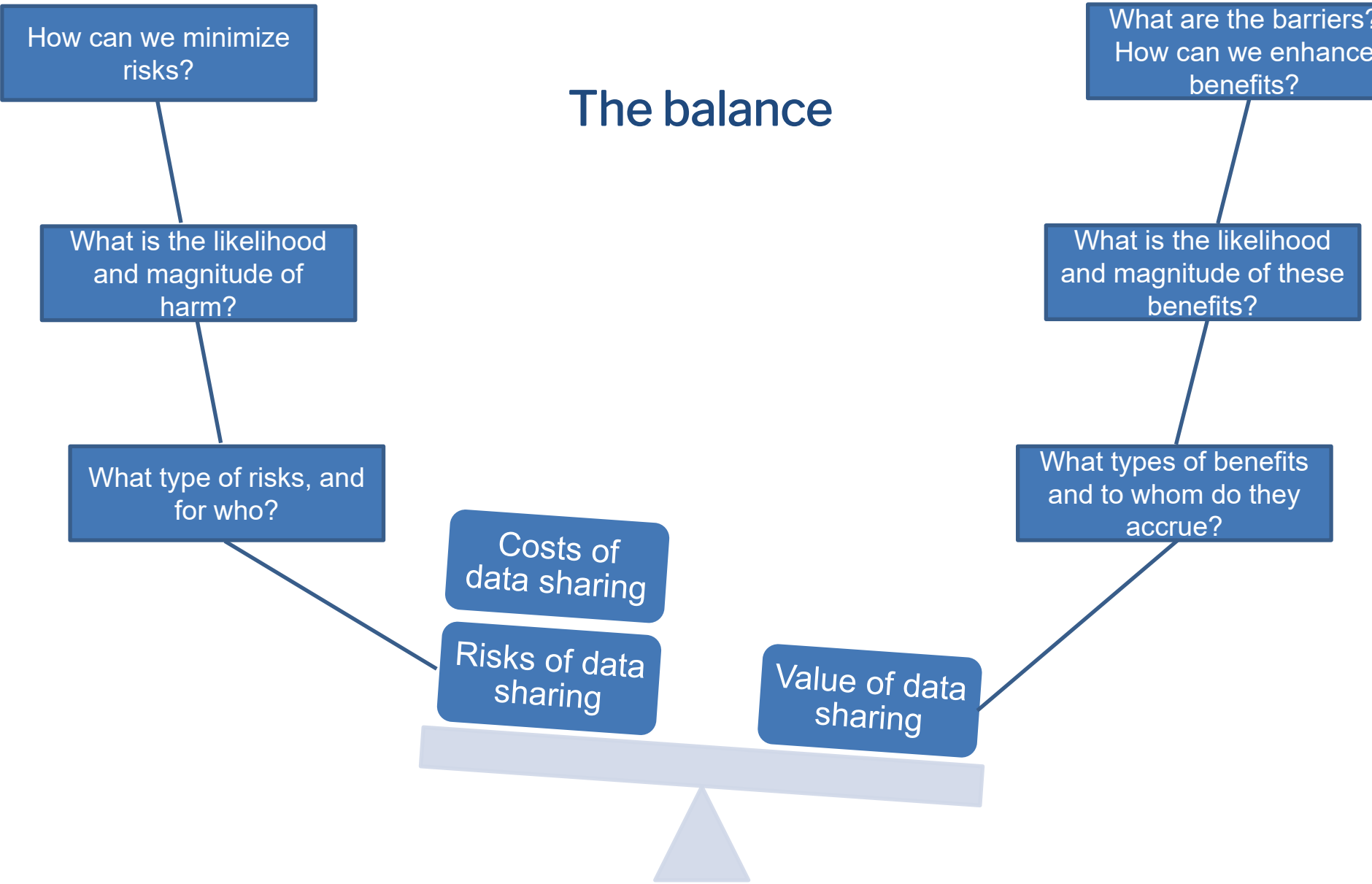
Learning objectives

- Consider the benefits and potential risks of data sharing;
- Appreciate that potential risks of sharing data differ across data types and contexts;
- Acquire insight into investigators' and research participants' views on data sharing;
- Argue why ethically appropriate protections for sharing data may be proportional to the potential risks of data sharing

Overview

- The value of data sharing
- Potential risks of data sharing
- Research participants' all things considered views
- Investigators' all things considered views
- Protections when sharing data

The balance


How can we minimize risks?
What is the likelihood and magnitude of harm?
What type of risks, and for who?
What are the barriers? How can we enhance benefits?
What is the likelihood and magnitude of these benefits?
What types of benefits and to whom do they accrue?
Costs of data sharing
Risks of data sharing
Value of data sharing

The value of data sharing

- Data long have been the cornerstone of science
- Improvements in technology, tools, and communications have made research data easier to manage, distribute, and reuse

The value of data sharing

The benefits of data sharing

Sharing data from research can be justified on scientific, economic and ethical grounds.

Data sharing can benefit:

- Future patients and society
- Researchers who shared the data
- Researchers who re-use the shared data
- Research participants

(Institute of Medicine, 2015; Taichman et al., 2016; Devriendt, et al., 2021, Wallis et al., 2013; Ohmann et al., 2017)

Benefits for society and future patients (1)

- Accelerate scientific advancement, improve safety and effectiveness of diagnostics and treatments
  - Further analysis of data (other hypotheses, meta analyses)
  - Increases transparency and accountability (e.g., verification)
  - Minimizes duplication of effort
  - Increases collaboration and interdisciplinary research
- Maximize value of research
  - Enhance value gained from participants' contributions
  - Enhance value gained from researcher's work
  - Greater returns on investment
- Improve trust in science

(Institute of Medicine, 2015; Taichman et al., 2016; Devriendt, et al., 2021, Chawinga & Zinn, 2019; Wallis et al., 2013)

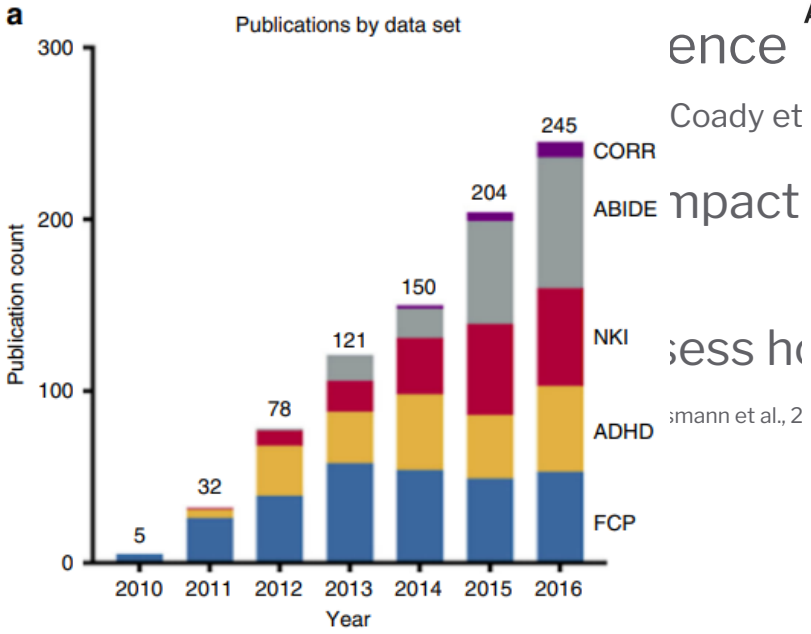
Benefits for society and future patients (2)

- Value of data sharing widely recognized (Shabani et all, 2014)
  - 75% neuroscience PIs thought there were no types of data in their field that were not worth the costs and efforts sharing (Hendriks et al., in prep)
- Yet, surprisingly little evidence on what happens once data are shared widely (Walker et al. 2011; Coady et al., 2017; Wallis, 2013)
  - Limited data to quantify impact or assess differential impact of types of data
  - Impact studies should assess how clinical trial data sharing influences knowledge generation (Mansmann et al., 2023)

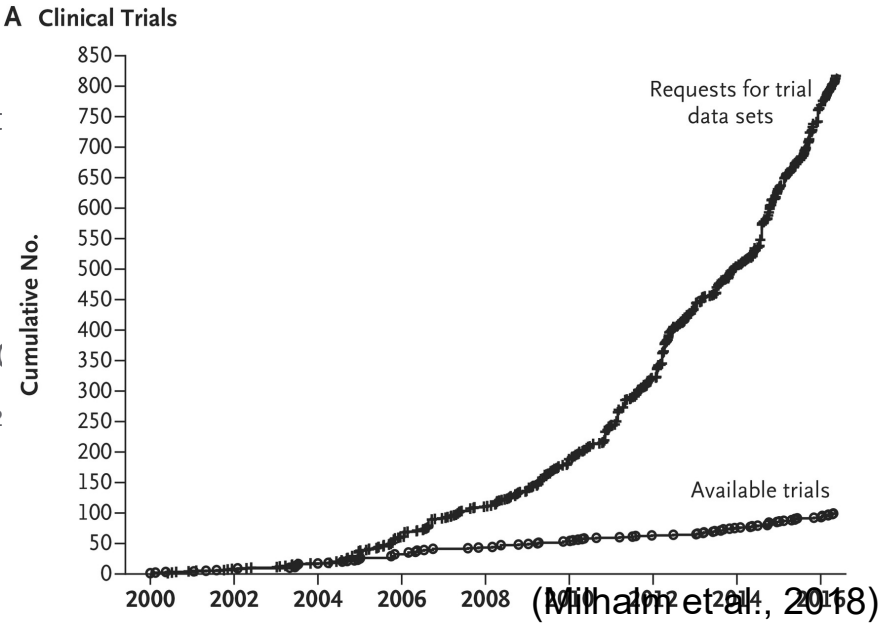
(Milhalm et al., 2018)

Benefits for society and future patients (2)

- Value of data sharing widely recognized (Shabani et all, 2014)
  - 75% neuroscience PIs thought there were no types of data in their field that were not worth the costs and efforts sharing (Hendriks et al., in prep)



Publications that used International Neuroimaging Data-sharing Initiative shared data (Milhalm et al., 2018)



(Milhalm et al., 2018)

Use of the NHLBI data repository (Coady et al., 2017)

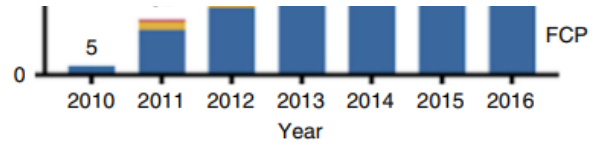
Benefits for society and future patients (2)

- Value of data sharing widely recognized (Shabani et all, 2014)
  - 75% neuroscience PIs thought there were no types of data in their field that were not worth the costs and efforts sharing (Hendriks et al., in prep)

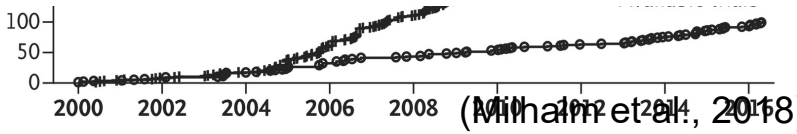
ence

Table 2 Quantifying the money saved through the reuse of data

Database	Cost/subject	Phenotyping	Phenotyping	Clinical	Population	Difficulty	No. of publications	No. of scans/ subject	$ Saved
		Minimal	Comprehensive	Low	Moderate	High			
FCP	$1000	x					308	1	101,003,000
ADHD-200	$2000–5000			x	x		210	1	526,275,000
NKI-RS	$3000		x				188	1	70,065,000
ABIDE	$5000–10,000				x	x	190	1	995,560,000
CoRR	$2000	x					17	2	70,065,000



Publications that used International Neuroimaging Data-sharing Initiative shared data (Milhalm et al., 2018)



(Milhalm et al., 2018)

Use of the NHLBI data repository (Coady et al., 2017)

Barriers to societal benefits

- 20-44% researchers self-report commonly using shared data (Hendriks et al., 2022; Tenopir et al., 2020; Science Staff, 2018)
  - 87% researchers self-report to be willing to use shared data (Tenopir et al., 2020)
- Difficult to access data
  - Data sharing platforms not consistently discoverable, searchable, and interoperable (National Research Council, 2015)
- Difficult to use data, e.g.,
  - limited standardization and norms for data acquisition, formatting, and description;
  - limited interpretation of data without understanding the context of data collection

  May result in erroneous secondary analyses

Among the top-5 barriers to data sharing according to neuroscientists (Hendriks et al., 2022)

Findable, Accessible, Interoperable, Reusable (FAIR) principles (Wilkinson et al., 2016)

Benefits for researchers who shared the data

- Association between articles with data in a repository and higher citation impact (Colavizza et al., 2020; Piwowar et al., 2007)
- Investigators' stated reasons to share data:
  - 42% of PIs: academic benefits and recognition (Rathi et al., 2012)
  - 69% of intramural NIH PIs: to collaborate with researcher who requested the data (Federer et al., 2015)

Primarily shared data upon request

	Attributions for data providers (Federer et al., 2015)
Acknowledgement	50%
Citation of the data	32%
Opportunity to collaborate	74% (coauthor)

Benefits for researchers who shared the data: barriers

- Few journal editors require data citations; granting agencies, tenure committees, and institutions may not reward data citations (National Research Council, 2012)
- >95% neuroscience investigators indicate that a lack of incentives is currently a barrier to data sharing (Hendriks et al., 2022)
- A lack of incentives is one of the main barriers to FAIR data sharing (Hughes et al., 2023)

Benefits for researchers who use the shared data

- Ability to do research without collecting original data; associated career-related benefits

Benefits for research participants

In as far as research participants value the value gained from their research contributions, data sharing can enhance their impact and may improve their sense that they contributed to something important.

- Participants who agreed to submit their data to dbGaP
  - 98% considered improving patient care and preventing or treating illness an important reason
  - 96% considered increasing knowledge for our society an important reason (Ludman et al., 2010)

A push towards data sharing

RECOMMENDATIONS

COMMISSION RECOMMENDATION (EU) 2018/790

of 25 April 2018

on access to and preservation of scientific information

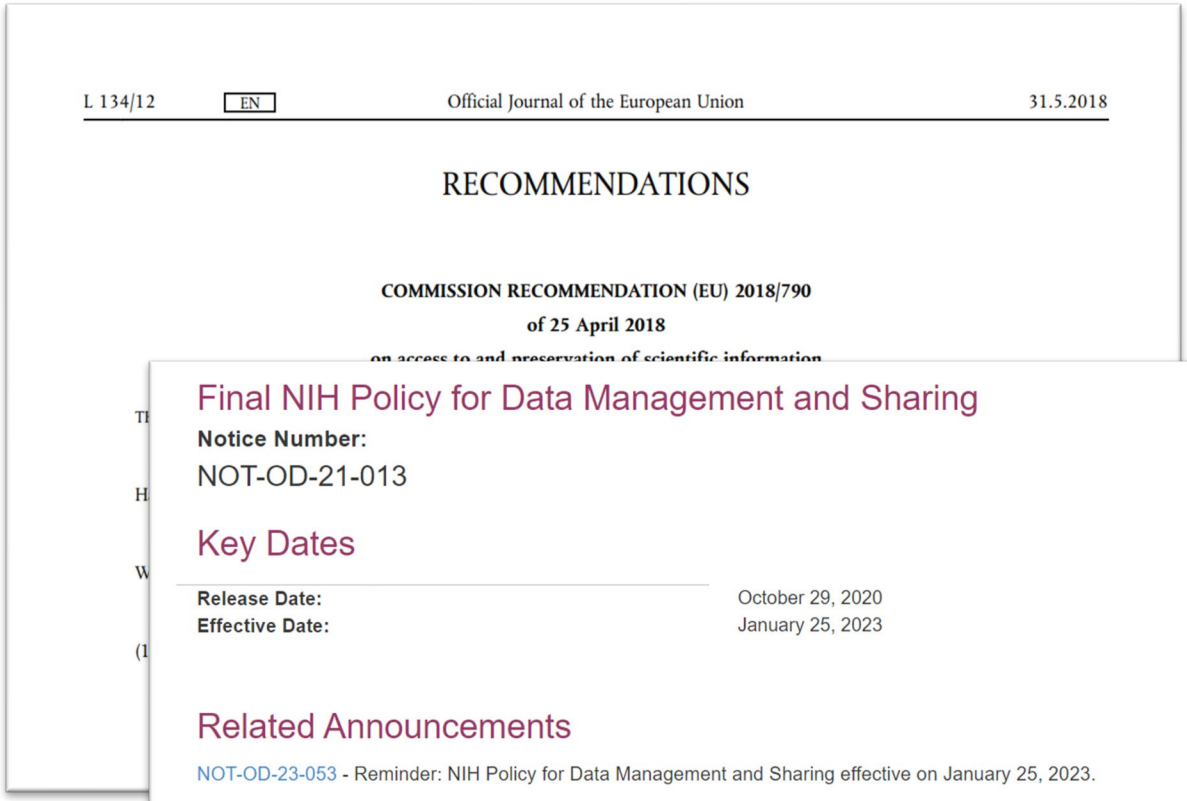
THE EUROPEAN COMMISSION,

Having regard to the Treaty on the Functioning of the European Union, and in particular Article 292 thereof,

Whereas:

(1) The European Commission adopted in July 2012 a scientific information package, consisting of the communication 'Towards better access to scientific information: Boosting the benefits of public investments in research' ([1]), and of Commission Recommendation 2012/417/EU ([2]). Recommendation 2012/417/EU states that the Commission will review the progress made across the Union to assess whether further action is needed to achieve the objectives laid down.

A push towards data sharing



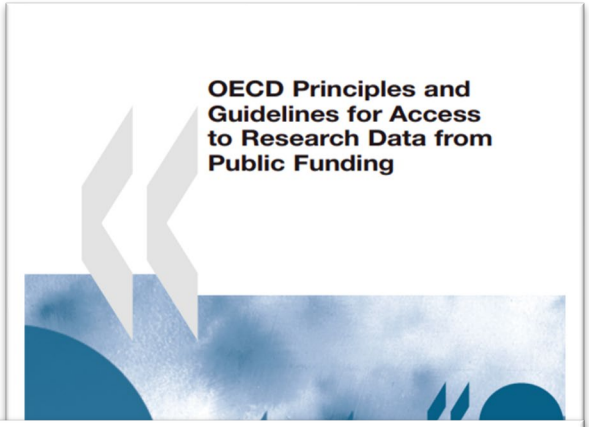
L 134/12 EN Official Journal of the European Union 31.5.2018

RECOMMENDATIONS

COMMISSION RECOMMENDATION (EU) 2018/790

of 25 April 2018

on access to and preservation of scientific information



Final NIH Policy for Data Management and Sharing

Notice Number:

NOT-OD-21-013

Key Dates

Release Date:	October 29, 2020
Effective Date:	January 25, 2023

Related Announcements

NOT-OD-23-053 - Reminder: NIH Policy for Data Management and Sharing effective on January 25, 2023.

NOT-OD-23-012 - Reminder: FORMS-H Grant Application Forms & Instructions Must be Used for Due Dates On or After January 25, 2023 - New Grant Application Instructions Now Available

NOT-CA-23-007 - Request for Information (RFI): Soliciting Input on the Use and Reuse of Cancer Metabolomics Data

NOT-OD-22-214 - Supplemental Information to the NIH Policy for Data Management and Sharing: Responsible Management and Sharing of American Indian/Alaska Native Participant

A push towards data sharing

A push towards data sharing

L 134/12 EN

National Science Foundation
WHERE DISCOVERIES BEGIN

Contact | Help

Search

Awards | Document Library | News | About NSF

...dget Finance & Award Managem... › Institution and Award Support

Email | Print | Share

...emination and Sharing of Research Results - NSF Data
...agement Plan Requirements

...TA SHARING POLICY

...tors are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the ...data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF ...rantees are expected to encourage and facilitate such sharing. See Proposal & Award Policies & Procedures Guide ... Chapter XI.D.4.

...n January 25, 2023.

...e Used for Due Dates On or After January 25, 2023 - New Grant Application Instructions Now

NOT-CA-23-007 - Request for Information (RFI): Soliciting Input on the Use and Reuse of Cancer Metabolomics Data

NOT-OD-22-214 - Supplemental Information to the NIH Policy for Data Management and Sharing: Responsible Management and Sharing of American Indian/Alaska Native Participant

The NEW ENGLAND JOURNAL of MEDICINE

EDITORIALS

Sharing Clinical Trial Data — A Proposal from the International Committee of Medical Journal Editors

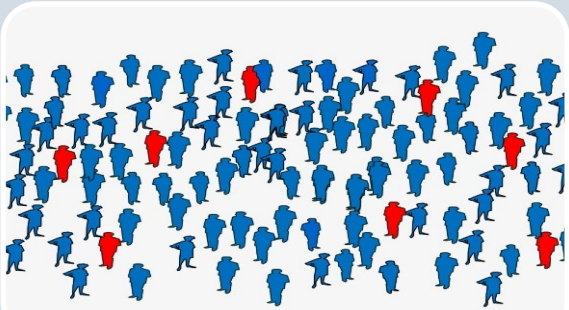
The International Committee of Medical Journal Editors (ICMJE) believes that there is an ethical obligation to responsibly share data generated by interventional clinical trials because participants have put themselves at risk. In a growing consensus, many funders around the world — foundations, government agencies, and industry — now mandate data sharing. Here we outline the ICMJE's proposed requirements to help meet this obligation. We encourage feedback on the proposed requirements. Anyone can provide feedback at www.icmje.org by 18 April 2016.

The ICMJE defines a clinical trial as any re-

their data are used. By changing the requirements of the manuscripts we will consider for publication in our journals, editors can help foster this endeavor. As editors, our direct influence is logically, and practically, limited to those data underpinning the results and analyses we publish in our journals.

The ICMJE also proposes to require that authors include a plan for data sharing as a component of clinical trial registration. This plan must include where the researchers will house the data and, if not in a public repository, the mechanism by which they will provide others

Potential harms of data sharing

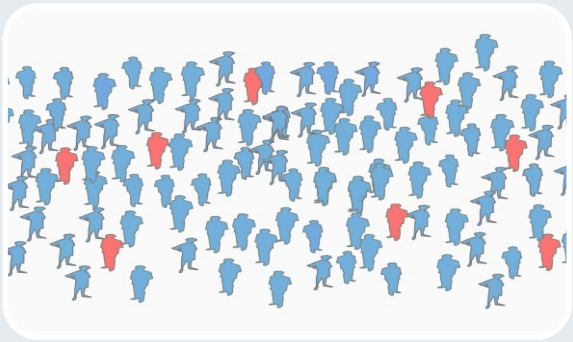


Individual harms

Community harms

Harms for research team, sponsor, funder, or institution

Individual harms



Community harms

Harms for research team, sponsor, funder, or institution

Potential individual harms

Identifiability

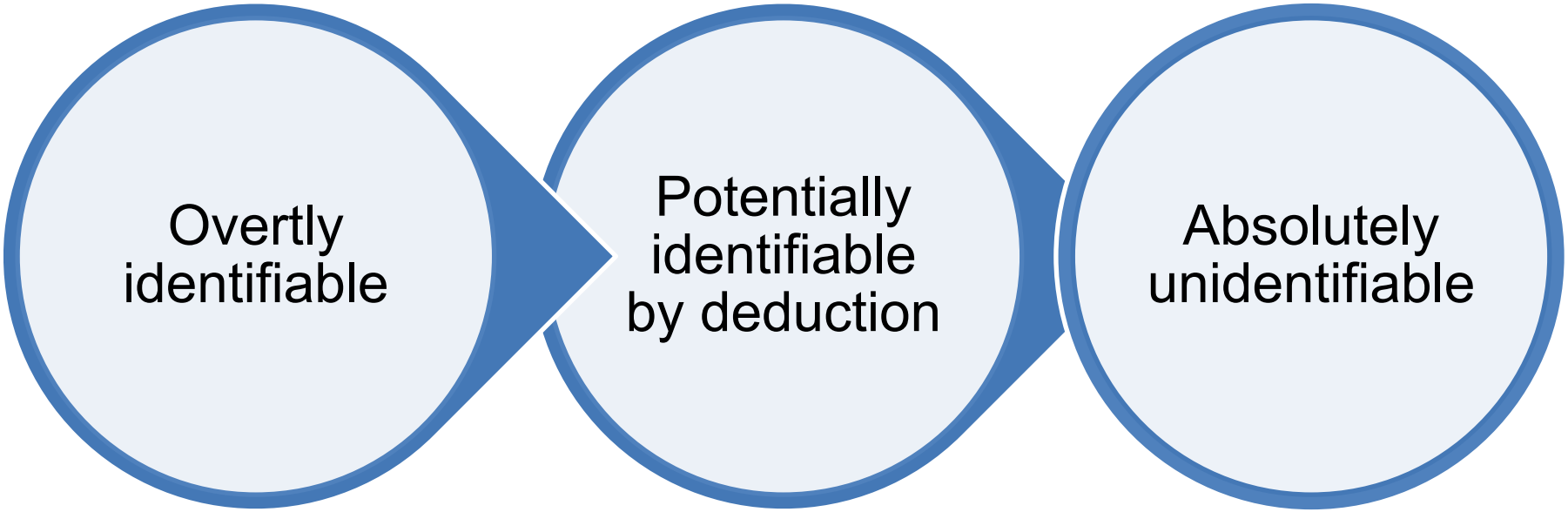
Harms for individual research participants

Potential individual harms

Identifiability

Harms for individual research participants

Identifiability



(Lowrance & Collins, 2007)

Identifiability and safeguards

Identifiability has implications for **ethics** and **oversight /regulations**

- Federal regulations for human subjects research (45CFR46)
- HIPAA (if covered entity or its associate)

- Potential implications for consent, ethics review, and other safeguards

(Lowrance & Collins, 2007)

Re-identification risk

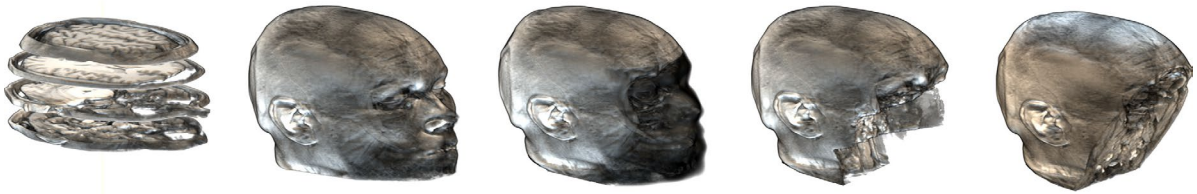
Risk of re-identification depends on:

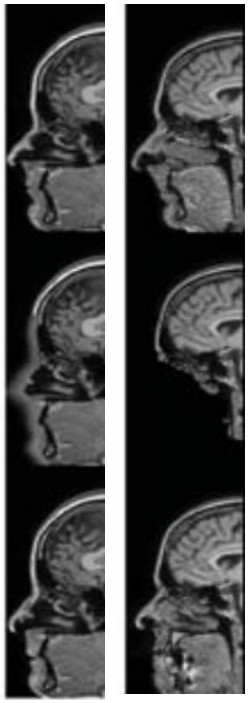
- Context in which data are released
- Type of data, e.g.,
  - Genomics data (Gymrek et al., 2013; Homer et al., 2008)
  - Certain types of brain data (Abramian and Eklund 2019; Bannier et al., 2021; Duan et al., 2020; Ravindra and Grama 2019; van de Ville et al., 2021)
  - Small N studies (van Mello et al., 2013)
- Additional information that might be combined with the shared data (e.g., from public databases)
- De-identification strategies used
  - Different standards, science continues to evolve

(Institute of Medicine, 2015; Lowrance & Collins, 2007)

Re-identification risk – example: MRI

- Faces are likely unique (Sheehan & Nachman, 2014)
  - 3D facial features can be reconstructed from MRI and matched to publicly accessible photos
  - Machine learning can "reface" single slice data (~60 to 75% success) (Abramian and Eklund 2019)





Original

Blurred/face-removed image

Reconstructed image

Bannier et al., 2021

De-identification challenges for investigators

- 39% of human genetics researchers have considerable discussion with their IRBs about procedures for protecting participants' personal information or samples (Edwards at al., 2011)
- 71% of neuroscience researchers think there is a risk that their de-identified data may be re-identified within the next 10 years (Hendriks et al., 2022)
  - 78% thought difficulty de-identifying data was a barrier to data sharing.

Likelihood of the risk

- Studies assess likelihood of re-identification of different types of data and de-identification strategies
- Limited data on how often research participants have been re-identified after data sharing outside of these published experiments
  - It is difficult to know whether there have been re-identification efforts

Likelihood of the risk

- Studies assess likelihood of re-identification of different types of data and de-identification strategies
- Published in final edited form as:
Technol Sci. 2017 ; 2017: .

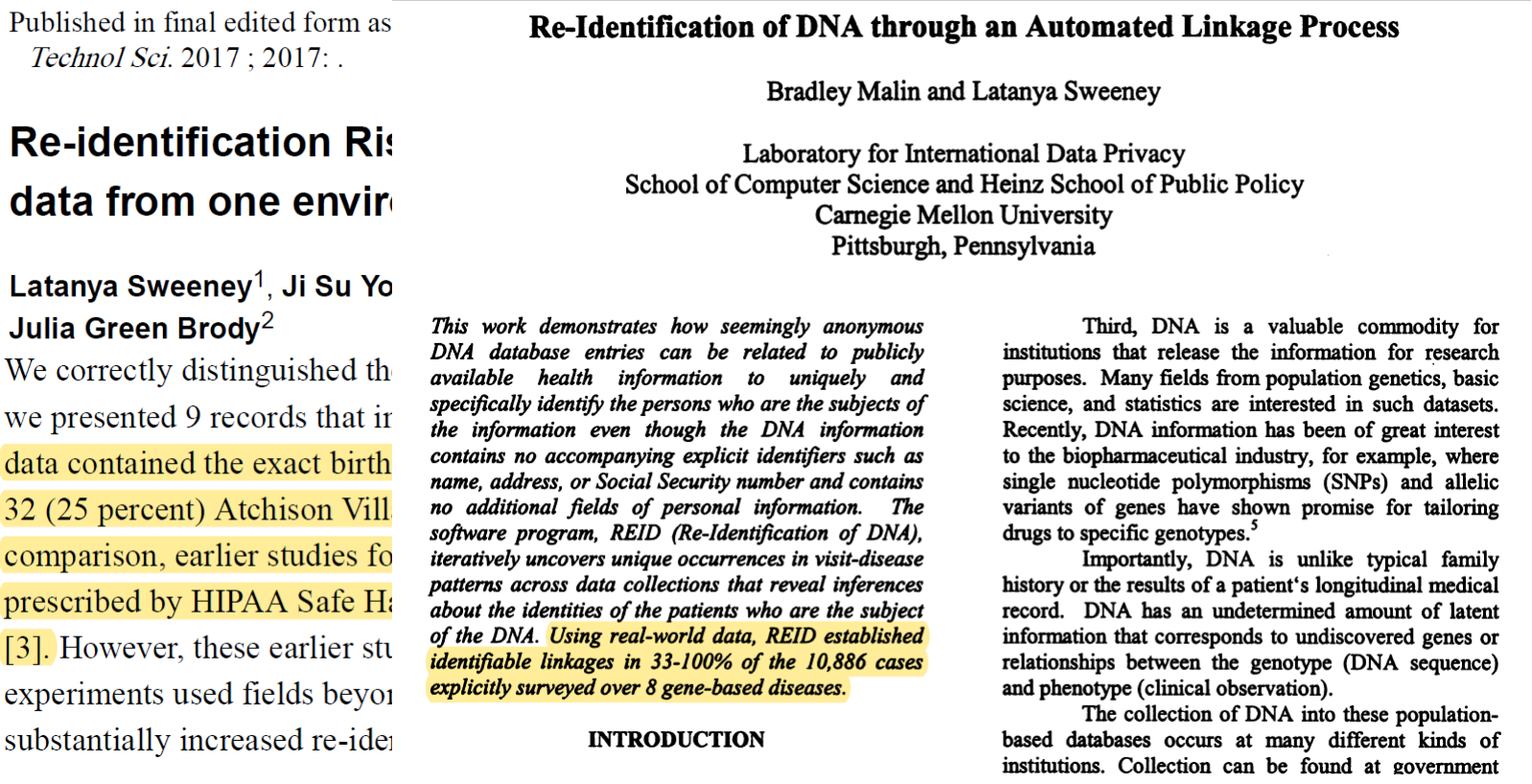
Re-identification Risks in HIPAA Safe Harbor Data: A study of data from one environmental health study

Latanya Sweeney[1], **Ji Su Yoo**[1], **Laura Perovich**[4], **Katherine E. Boronow**[2], **Phil Brown**[3], and **Julia Green Brody**[2]

We correctly distinguished the 10 records from Bolinas and 32 records from Atchison Village, and we presented 9 records that included the 8 correct records from Liberty Village. When the redacted data contained the exact birth year, as allowed by HIPAA Safe Harbor, we correctly identified 8 of 32 (25 percent) Atchison Village participants by name and 9 of 32 (28 percent) by address. In comparison, earlier studies found unique re-identification rates in data that adhered to the level prescribed by HIPAA Safe Harbor to be much lower, namely 0.013 percent [2] and 0.04 percent [3]. However, these earlier studies relied solely on demographic fields for re-identification. Our experiments used fields beyond demographics (e.g., housing characteristics), and by doing so, substantially increased re-identification risk in data compliant with HIPAA Safe Harbor. Even in

Likelihood of the risk

- Studies assess likelihood of re-identification of different types of data and de-identification strategies



Published in final edited form as
Technol Sci. 2017 ; 2017: .

Re-identification Ris
data from one envir

Latanya Sweeney[1], **Ji Su Yo**
Julia Green Brody[2]

We correctly distinguished th
we presented 9 records that ir
data contained the exact birth
32 (25 percent) Atchison Vill
comparison, earlier studies fo
prescribed by HIPAA Safe Ha
[3]. However, these earlier stu
experiments used fields beyo
substantially increased re-ide

Re-Identification of DNA through an Automated Linkage Process

Bradley Malin and Latanya Sweeney

Laboratory for International Data Privacy
School of Computer Science and Heinz School of Public Policy
Carnegie Mellon University
Pittsburgh, Pennsylvania

This work demonstrates how seemingly anonymous DNA database entries can be related to publicly available health information to uniquely and specifically identify the persons who are the subjects of the information even though the DNA information contains no accompanying explicit identifiers such as name, address, or Social Security number and contains no additional fields of personal information. The software program, REID (Re-Identification of DNA), iteratively uncovers unique occurrences in visit-disease patterns across data collections that reveal inferences about the identities of the patients who are the subject of the DNA. Using real-world data, REID established identifiable linkages in 33-100% of the 10,886 cases explicitly surveyed over 8 gene-based diseases.

INTRODUCTION

Third, DNA is a valuable commodity for institutions that release the information for research purposes. Many fields from population genetics, basic science, and statistics are interested in such datasets. Recently, DNA information has been of great interest to the biopharmaceutical industry, for example, where single nucleotide polymorphisms (SNPs) and allelic variants of genes have shown promise for tailoring drugs to specific genotypes.[5]

Importantly, DNA is unlike typical family history or the results of a patient's longitudinal medical record. DNA has an undetermined amount of latent information that corresponds to undiscovered genes or relationships between the genotype (DNA sequence) and phenotype (clinical observation).

The collection of DNA into these population-based databases occurs at many different kinds of institutions. Collection can be found at government

Potential individual harms

Identifiability

Harms for individual research participants

Types of potential individual harms

- Privacy
- Psychological harms
- Social and economic harms
  - Discrimination
  - Reputational harms
- Legal harms
- Personal and financial security
- Exposure to unwanted attention/service offers
- Dignity
- Use of data in ways inconsistent with participants' values

(Fuchs, 2006, Ienca and Haselager, 2016, Illes and Racine, 2005, Kolber, 2007, Lavazza, 2018, Rommelfanger, et al., 2018, Panel, 2006, Räikkä, 2010; Arstila and Scott, 2011, Bublitz, 2011, Räikkä, 2010, Lavazza, 2018, Tunick, 2017, Bonaci, et al., 2014, Bonaci, et al., 2015,, Ienca, et al., 2018; Scott et al., 2018)

Privacy

- Privacy has proved notoriously difficult to define (Finn, et al., 2013, Lever, 2011, Matthews, 2015, Moore, 2008, Rommelfanger, et al., 2018)
  - The **state or condition** of being let alone, free from being observed or disturbed by other people

  - The ability to **control** the extent, timing, and circumstances of sharing oneself (physically, behaviorally, intellectually) with others

Investigators' concerns about risks for participants

- 29-91% of investigators are concerned about risks to participants (Tan et al., 2021; Hendriks et al., 2022; Rathi et al., 2012)
  - Most common reason not to share data (Tan et al., 2021)
  - Investigators who collect data they think may be re-identifiable or data from vulnerable groups are more concerned about risk to participants. (Hendriks et al., 2022)
- Even if they *also* think harms to research participants are unlikely (Hendriks et al., 2022)

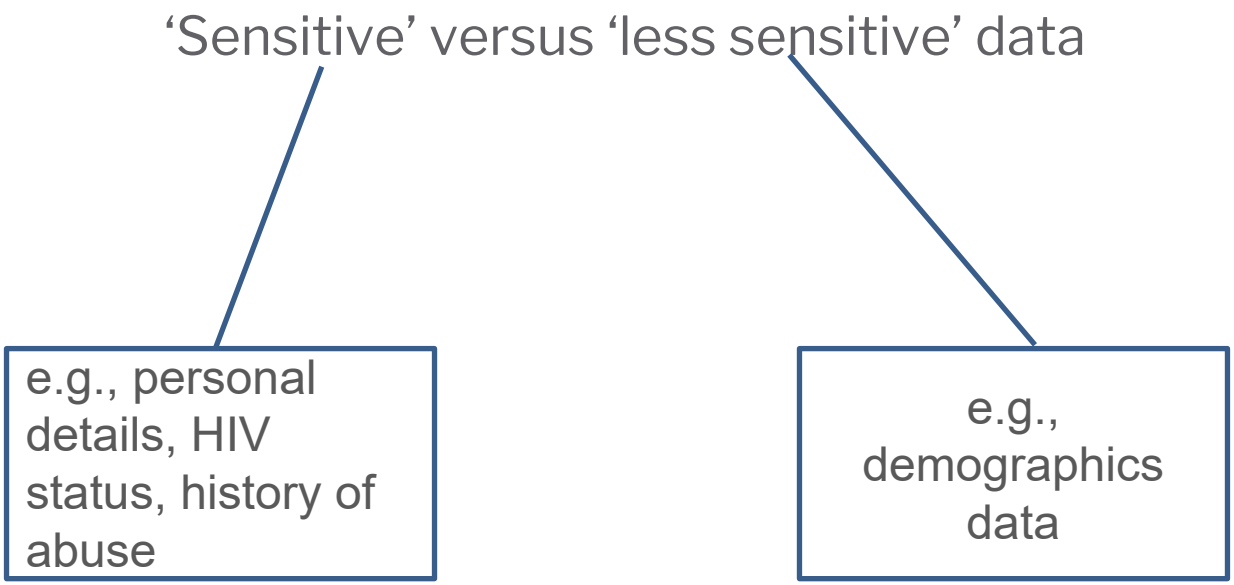
Participants' concerns about risks for participants (1)

- Re-identification and privacy:
  - Reported on in most studies in a systematic review of participants' attitudes towards data sharing (Howe et a., 2018)
  - 33-61% of participants or potential participants concerned (Ludman et al., 2010; Sanderson et al., 2017; Mello et al., 2018)
- Stigma and discrimination (Howe et a., 2018; Shabani et al., 2014)
  - 49% of trial participants are concerned (Mello et al., 2018)
- Embarrassment
  - 43% of trial participants concerned about embarrassment (Mello et al., 2018)

Participants' concerns about risks for participants (2)

- Informational security
  - 64% of trial participants concerned their information may be stolen (Mello et al., 2018)
  - "Mundane concerns" such as telemarketing (Howe et al., 2018)
- "Repercussions"
  - 37% of the US public concerned data from a genetic study could be used against them (Kaufman et al., 2009)
  - 79% potential participants want to know if data might be used by insurance companies (Sanderson et al., 2017)
  - e.g., being reported to social services (Howe et a., 2018)

Participants' concerns about risks for participants (3)



The potential sensitivity of data was related to its use

(Howe et a., 2018)

Family

...

Quantifying risks

Limited evidence that allows for quantifying

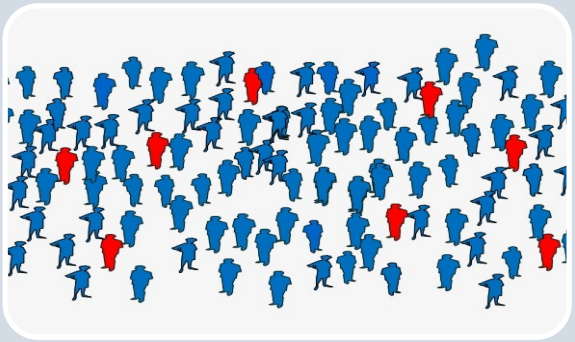
- The likelihood of these risks
- The severity of the harms
- Distribution of any harms

Genetic Information Non-Discrimination Act Charges (Charges filed with EEOC) by FY

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Receipts	201	245	280	333	333	257	238	206	220	209	440	242
Merit Resolutions	7	36	77	63	57	50	38	55	45	36	81	47
	12.5%	17.1%	24.1%	21.4%	18.5%	19.5%	15.3%	20.6%	15.6%	12.7%	30.8%	19.8%

Source: https://www.eeoc.gov/data/genetic-information-non-discrimination-act-charges-charges-filed-eeoc-includes-concurrent

Individual harms



Community harms

Harms for research team, sponsor, or institution

Potential societal and group harms

Unskilled analysts, market competitors, or others with strong private agendas publish poorly conducted analyses

- Could mislead clinicians and patients, negatively affect patient care
- Could undermine trust in research
- Could otherwise harm public health
- Could impinge on health and non-health related interests of groups (e.g., stigma and discrimination; dignity; loss of services; or uses that the community finds objectionable)
- *(Could lead to expensive lawsuits without merit)*
- *(Could cause reputational harms for investigators)*
- *(Responding could be a high burden for investigators)*

(Scott et al., 2018; Ross et al., 2018; Sabatello et al. 2022; Institute of Medicine 2015; Mello 2013)

Some examples

Journal of Health & Biomedical Law, VI (2010): 175-225

Suffolk University Law School

Lessons from Havasupai Tribe v. Arizona State University Board of Regents: Recognizing Group, Cultural, and Dignitary Harms as Legitimate Risks Warranting Integration into Research Practice

Katherine Drabiak-Syed*

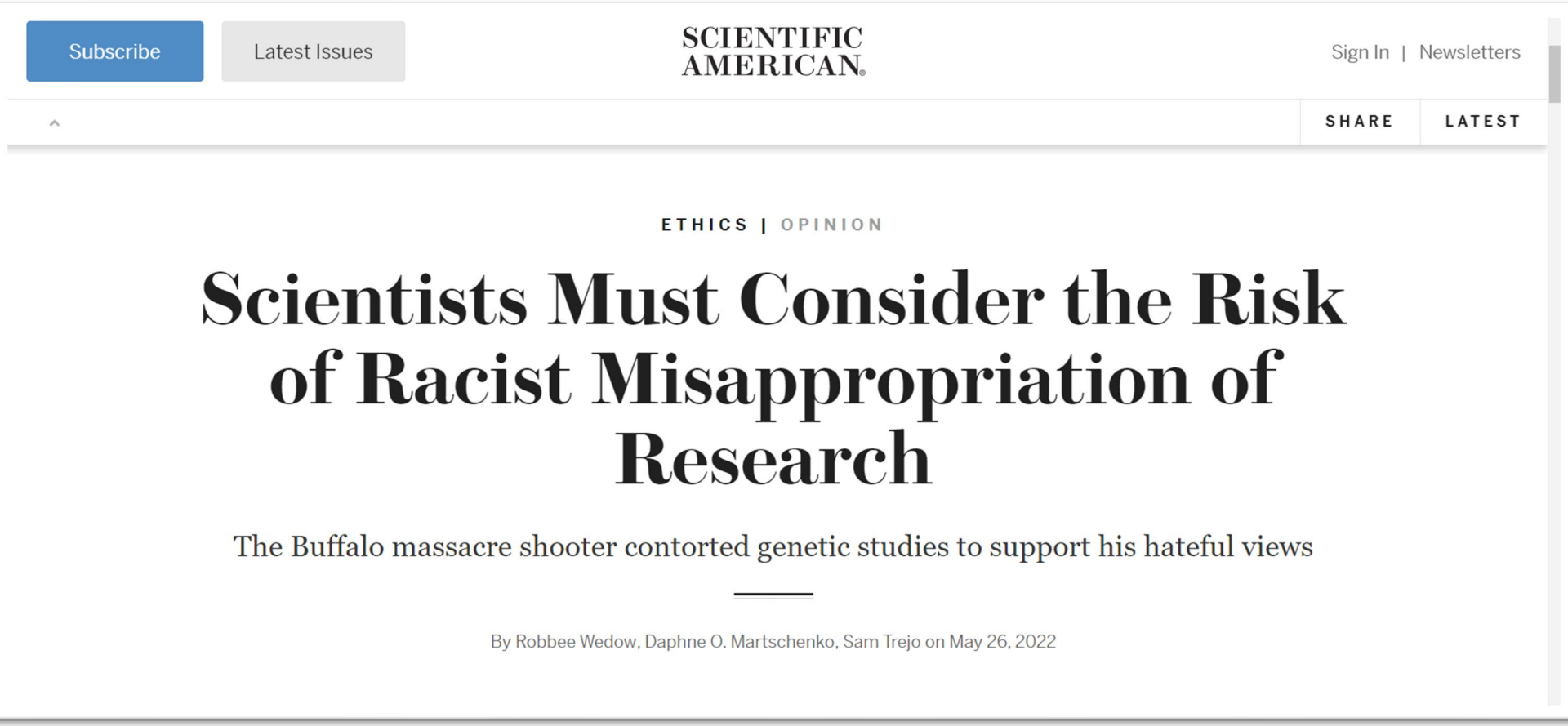
Some examples

AI recognition of patient race in medical imaging: a modelling study

Judy Wawira Gichoya, Imon Banerjee, Ananth Reddy Bhimireddy, John L Burns, Leo Anthony Celi, Li-Ching Chen, Ramon Correa, Natalie Dullerud, Marzyeh Ghassemi, Shih-Cheng Huang, Po-Chih Kuo, Matthew P Lungren, Lyle J Palmer, Brandon J Price, Saptarshi Purkayastha, Ayis T Pyrros, Lauren Oakden-Rayner, Chima Okechukwu, Laleh Seyyed-Kalantari, Hari Trivedi, Ryan Wang, Zachary Zaiman, Haoran Zhang

Interpretation The results from our study emphasise that the ability of AI deep learning models to predict self-reported race is itself not the issue of importance. However, our finding that AI can accurately predict self-reported race, even from corrupted, cropped, and noised medical images, often when clinical experts cannot, creates an enormous risk for all model deployments in medical imaging.

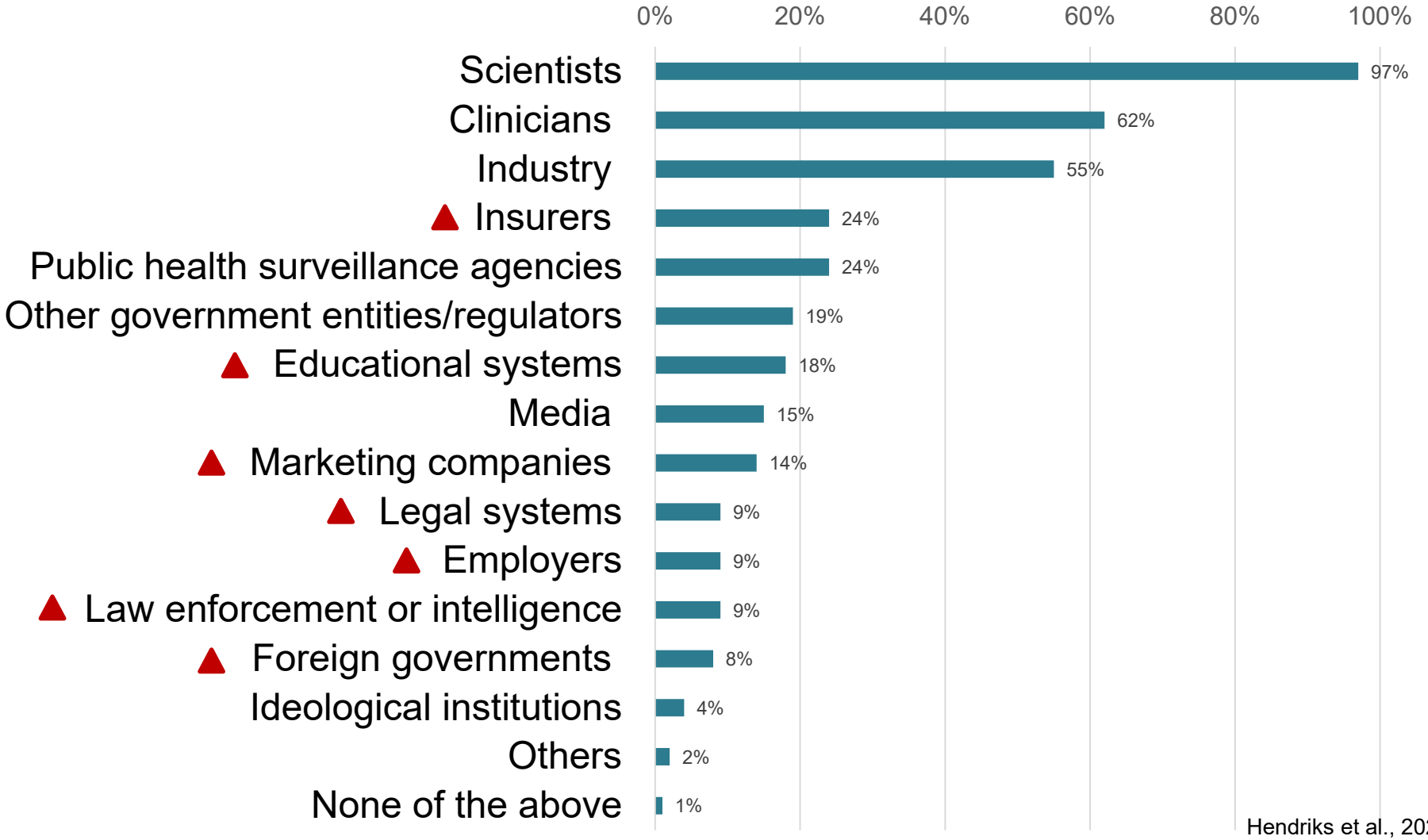
Some examples



Investigators' concerns

- Although most neuroscience investigators consider misuse unlikely, the majority were concerned about potential group harms from their own data (63%) (Hendriks et al., 2022)
  - 85% considers the risk of misuse a barrier to data sharing
- Inappropriate data use was the most common concern among clinical trialists about data sharing through repositories (65%) (Rathi et al., 2012)
- 75% of investigators think research data may be used in "different ways than intended" (Tenopir et al., 2020)
- 79% of investigators think data may be misinterpreted (Tenopir et al., 2020)

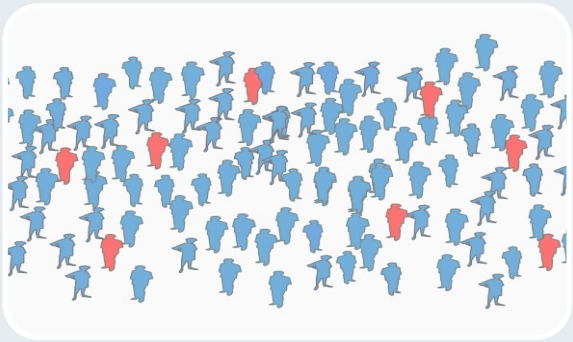
Who neuroscience investigators think would be interested in their data

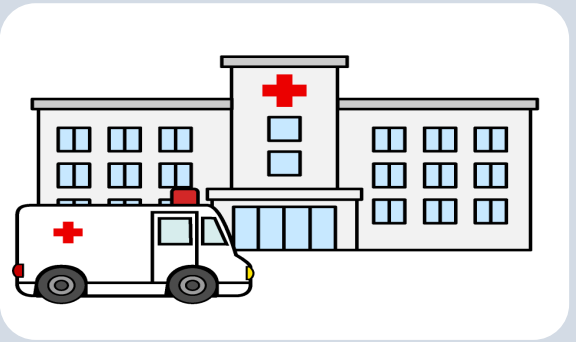


Hendriks et al., 2022

Participants' concerns

- 52% of participants who agreed to submit data to dbGaP thought misuse concerns were important (Ludman et al., 2010)
- 84% of potential participants would want to know who makes sure that their data is used "in the right way" (Sanderson et al., 2017)
- Re-use that's inconsistent with participants' values (Howe et al., 2018; Shabani et al., 2014)
  - 36-66% of trial participants concerned about re-use for marketing (Mello et al., 2018; Sanderson et al., 2017)
  - 37% of participants concerned about secondary research they wouldn't have wanted (Sanderson et al., 2017)
  - Commodification of data (Spector-Bagdady et al., 2020)





Individual harms

Group harms

Harms for research team, sponsor, funder, or institution

Sponsors, research institutions, and research teams

Have legitimate interests which could be undermined through data sharing, depending on the matter at which sharing occurs:

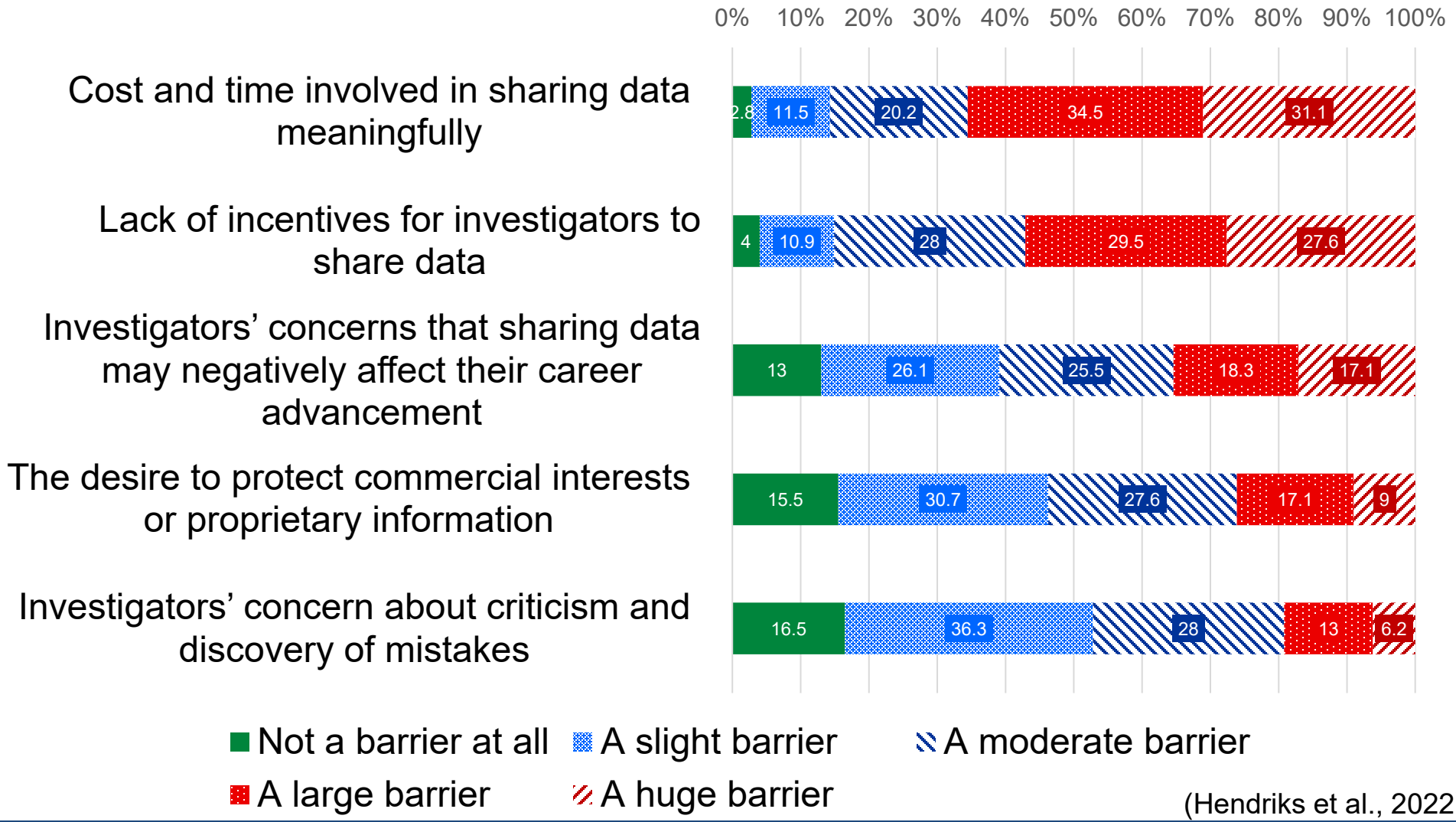
- Intellectual property; commercially confidential information
- Repercussions from invalid secondary analyses
- Career advancement, intellectual capital, and professional recognition
- Undue costs of data sharing itself
- Legal risks related to data misuse

(Institute of Medicine 2015; Scott et al., 2018)

Investigators' concerns

- Commercially confidential information
  - 91% of investigators concerned (Tan et al., 2021)
- Career advancement, intellectual capital, and professional recognition
  - 91-93% of investigators concerned (Tan et al., 2021)
- Undue costs of data sharing
  - 96% of investigators concerned about time and effort (Tan et al., 2021)
  - 83% of investigators concerned about costs (Tan et al., 2021)
- 41% of clinical trialists concerned about investigator and funder interests (Rathi 2012)

Neuroscience investigators: barriers to data sharing



(Hendriks et al., 2022)

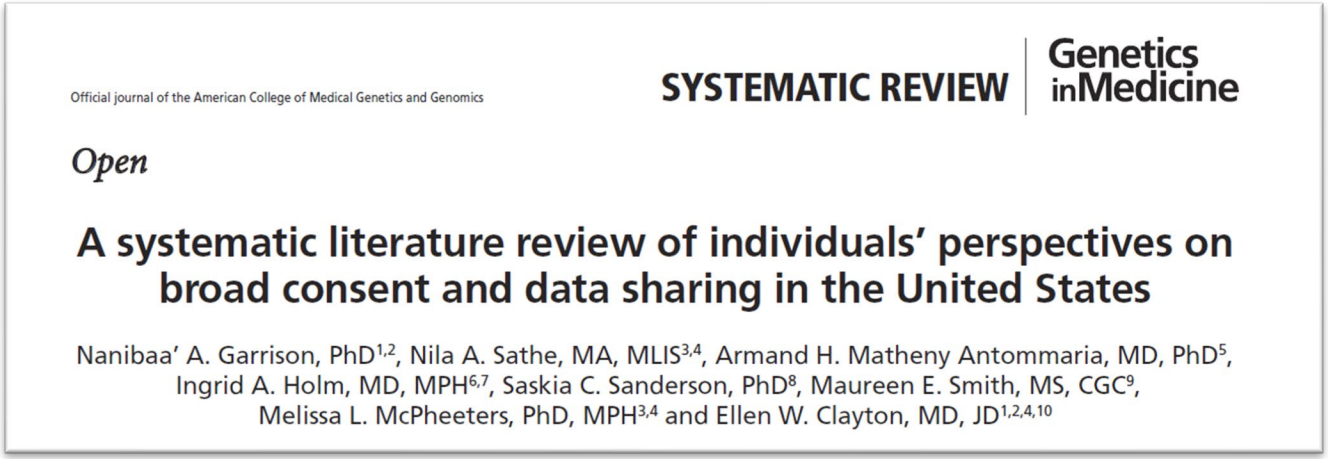
Sponsors, research institutions, and research teams

Sharing of clinical trial data needs to be carried out in a way that maintains incentives to develop new therapies and carry out future clinical trials
(Institute of Medicine 2015)

Research participants' all things considered views

Empirical research on research participants' views

- Sys Review (2015): 48 studies on participant views on data sharing for biobank research (Garrison et al., 2015).
  - Studies included a total of 35,969 individuals.
- … And more since



Official journal of the American College of Medical Genetics and Genomics

SYSTEMATIC REVIEW | Genetics inMedicine

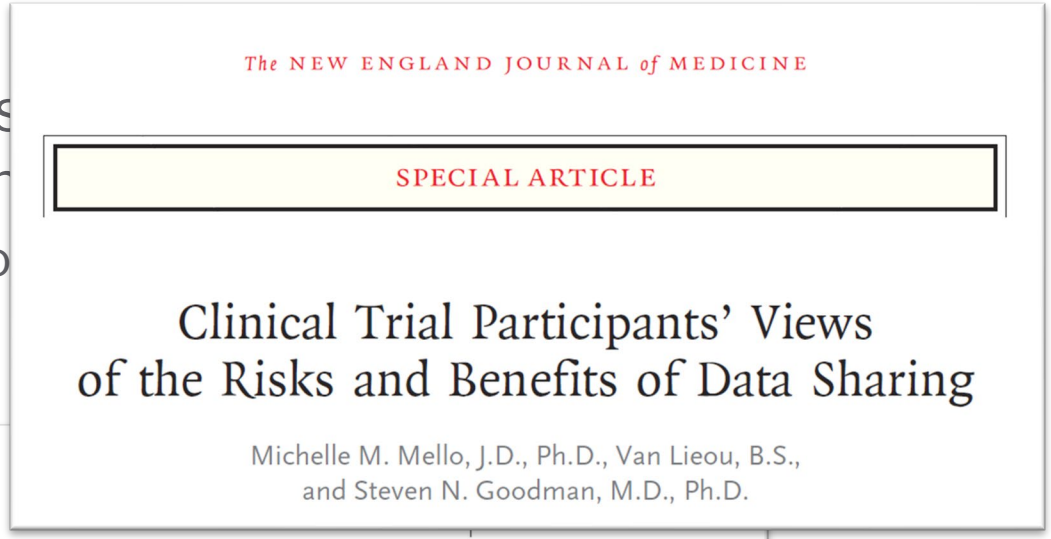
Open

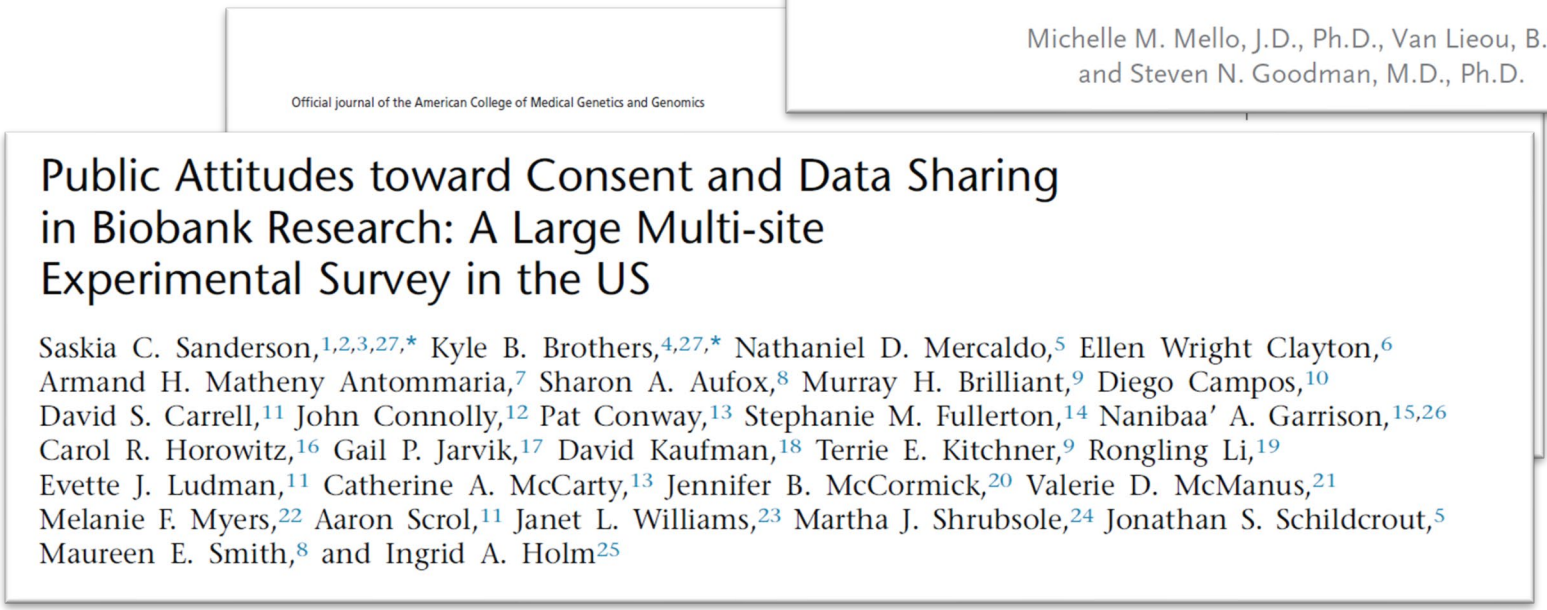
A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States

Nanibaa' A. Garrison, PhD[1,2], Nila A. Sathe, MA, MLIS[3,4], Armand H. Matheny Antommaria, MD, PhD[5], Ingrid A. Holm, MD, MPH[6,7], Saskia C. Sanderson, PhD[8], Maureen E. Smith, MS, CGC[9], Melissa L. McPheeters, PhD, MPH[3,4] and Ellen W. Clayton, MD, JD[1,2,4,10]

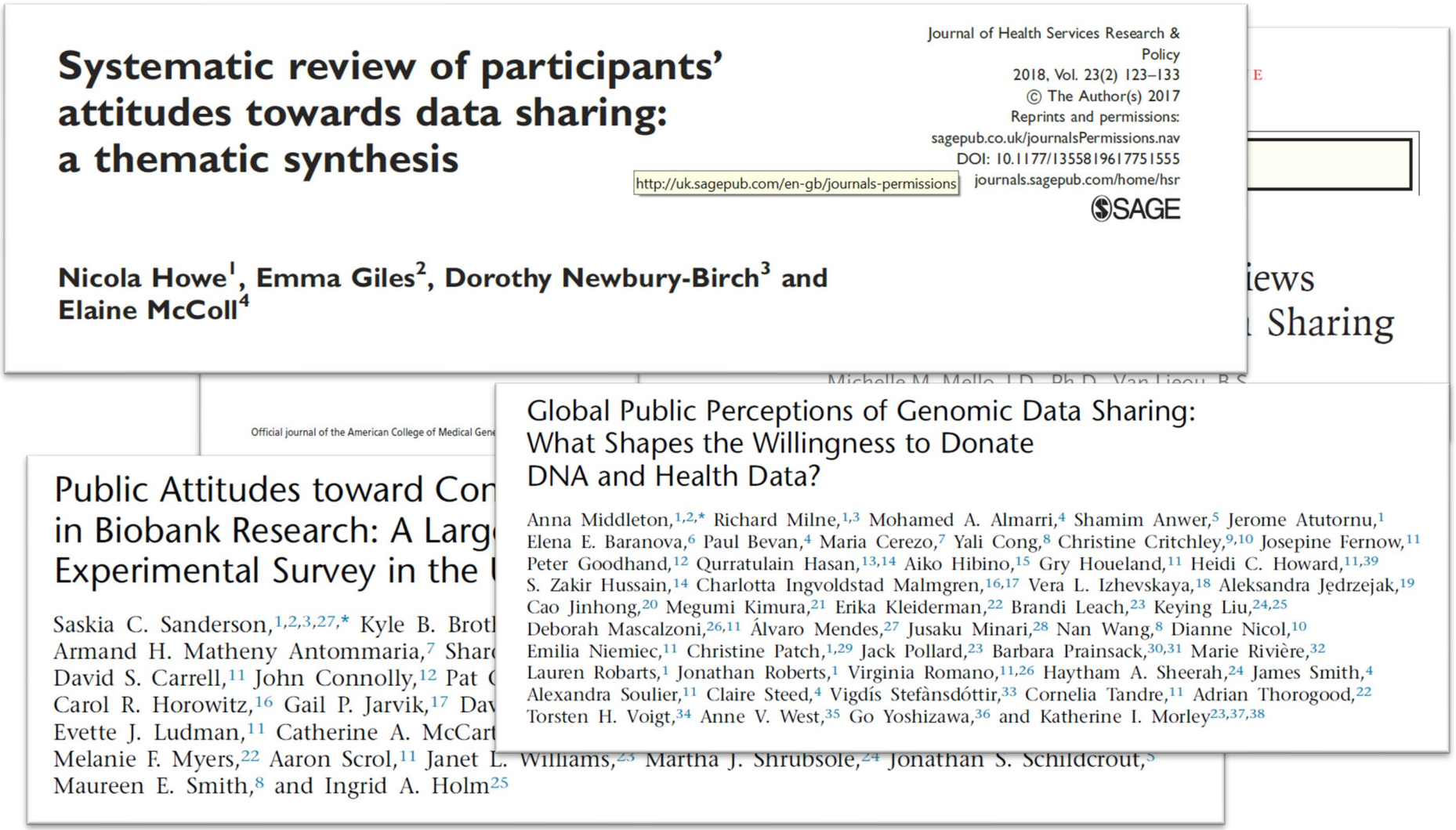
Empirical research on research participants' views

- Sys Review (2015): 48 s
  data sharing for biobar
  - Studies included a total o
- … And more since





Empirical research on research participants' views



Systematic review of participants' attitudes towards data sharing: a thematic synthesis

Journal of Health Services Research & Policy
2018, Vol. 23(2) 123–133

DOI: 10.1177/1355819617751555
journals.sagepub.com/home/hsr
SAGE

Nicola Howe[1], Emma Giles[2], Dorothy Newbury-Birch[3] and Elaine McColl[4]

Global Public Perceptions of Genomic Data Sharing: What Shapes the Willingness to Donate DNA and Health Data?

Anna Middleton,[1,2,*] Richard Milne,[1,3] Mohamed A. Almarri,[4] Shamim Anwer,[5] Jerome Atutornu,[1] Elena E. Baranova,[6] Paul Bevan,[4] Maria Cerezo,[7] Yali Cong,[8] Christine Critchley,[9,10] Josepine Fernow,[11] Peter Goodhand,[12] Qurratulain Hasan,[13,14] Aiko Hibino,[15] Gry Houeland,[11] Heidi C. Howard,[11,39] S. Zakir Hussain,[14] Charlotta Ingvoldstad Malmgren,[16,17] Vera L. Izhevskaya,[18] Aleksandra Jędrzejak,[19] Cao Jinhong,[20] Megumi Kimura,[21] Erika Kleiderman,[22] Brandi Leach,[23] Keying Liu,[24,25] Deborah Mascalzoni,[26,11] Álvaro Mendes,[27] Jusaku Minari,[28] Nan Wang,[8] Dianne Nicol,[10] Emilia Niemiec,[11] Christine Patch,[1,29] Jack Pollard,[23] Barbara Prainsack,[30,31] Marie Rivière,[32] Lauren Robarts,[1] Jonathan Roberts,[1] Virginia Romano,[11,26] Haytham A. Sheerah,[24] James Smith,[4] Alexandra Soulier,[11] Claire Steed,[4] Vigdís Stefànsdóttir,[33] Cornelia Tandre,[11] Adrian Thorogood,[22] Torsten H. Voigt,[34] Anne V. West,[35] Go Yoshizawa,[36] and Katherine I. Morley[23,37,38]

Official journal of the American College of Medical Gene

Public Attitudes toward Con
in Biobank Research: A Larg
Experimental Survey in the

Saskia C. Sanderson,[1,2,3,27,*] Kyle B. Brotl
Armand H. Matheny Antommaria,[7] Shar
David S. Carrell,[11] John Connolly,[12] Pat
Carol R. Horowitz,[16] Gail P. Jarvik,[17] Dav
Evette J. Ludman,[11] Catherine A. McCart
Melanie F. Myers,[22] Aaron Scrol,[11] Janet L. Williams,[23] Martha J. Shrubsole,[24] Jonathan S. Schildcrout,[3]
Maureen E. Smith,[8] and Ingrid A. Holm[25]

Participants' support of data sharing

- 82% research participants think the benefits of data sharing outweighed the negative aspects
  - 8% felt the negative aspects outweighed the benefits (Mello et al., 2018)
- Hypothetical choices:
  - 97% research participants supported their data being shared (Shah et al., 2019)
  - 60-88% willingness to broadly share samples/data through biobank (Garrison et al., 2015; Sanderson et al., 2017)
- Actual choices re: biobank participation with broad consent:
  - 87% of research participants at the NIH CC authorized all future research (n= 1,298) (Chen et al. 2005)
  - 85% of participants agreed to DNA specimen in a national repository (n=4,480) (Mcquillan et al., 2003)

Factors that may influence participants' willingness (1)

- Type of data
  - Degrees of 'sensitivity' (Bell et al., 2014; Middelton et al, 2020; McGuire et al., 2011)
- Access model
  - Controlled versus open access (Sanderson et a., 2017; McGuire et al., 2011; Haga & O'Daniel 2010; Kaufman et al., 2009)
- Type of secondary user
  - Academic researcher in same geographic region -- for-profit organizations -- insurers and employers (Shah et al., 2019; Garrison et al., 2015; Shabani et al., 2014; Mello et al., 2018; Kaufman et al., 2009)
- Type of secondary use
  - Biomedical research vs lawsuits (Sanderson et a., 2017)

Factors that may influence participants' willingness (2)

- How well-informed they are about data sharing (McGuire et al., 2011)
- Participants' demographics
  - E.g., self-identified white, non-Hispanics more willing to share . (Sanderson et a., 2017; McGuire et al., 2011; Kaufman et al., 2009)
- Participants' attitudes
  - E.g., more trust in related institutions (Sanderson et al., 2017; Mello et al., 2018; Kaufman et al., 2009)

Consider engaging with communities that are uncomfortable with sharing their data

Investigators' all things considered views and behaviors

Stated willingness of investigators to share data

- 77-88% of researchers support data sharing (Rathi et al., 2012; Tan et al., 2021; Hendriks et al., 2022; Zhu et al., 2019)
  - Consistently more support for the *idea* of data sharing than actual practices of data sharing (Thoegersen and Pia Borlund 2021)

- Willingness to share data depends on how "broadly" (Tenopir et al., 2020)
  - 87% willing to share data with researchers
  - 77% willing to share some data open access
  - 45% willing to share all data open access

- Younger researchers feel more favorably toward data sharing and reuse, yet make less of their data available (Tenopir et al, 2015)

The prevalence of data sharing

- 71-91% of researchers self-report to have ever shared data (Federer et al., 2015;Tenopir et al., 2015; Hendriks et al., 2022)
- 33-39% of researchers self-report to have shared data through archives (Federer et al., 2015; Tenopir et al., 2015)
  - 3-39% through open access archives (Thoegersen and Pia Borlund 2021)
- Trend towards increased data sharing
  - Increase in self-reported data sharing behaviors between 2009-2014 (Tenopir et al., 2015)
- Data availability across journals and disciplines: 9-76% (Tedersoo et al., 2021)
- 27–59% of data requests to authors are successful (Tedersoo et al., 2021)

Protections

Types of protections

- Data security
- Informed consent
- Controlled access
- Governance / regulations/ policy

Data security

Informed consent

- Participants want to give consent re: de-identified data sharing
  - Important to up to 90% of participants (Ludman et al., 2010)
- Ethically, many have argued that studies that will enroll participants prospectively, the informed consent procedure can –-and should – address data sharing
- Different models for what kind of consent (broad consent, categorical consent, study-by-study)
  - Many participants support broad consent, but individuals do differ in terms of their preferred type of consent for biobank research (Garrison et al., 2015; Howe et a., 2018; Sanderson et a., 2017; Kaufman et al., 2009; Warner et al., 2018)

Informed consent: understanding and disclosure

- Disclosures about data sharing ≠ understanding (Spector-Bagdady et al., 2020; Valle-Mansilla et al., 2010)
  - —82% remembered giving informed consent that their samples could be used for future research (Valle-Mansilla et al., 2010)

- Participants change what they consent to after more information
  - 32% change what they consent to after debriefing (McGuire et al., 2011)
  - 18% change what they consent to after informational video (Riggs et al., 2019)
  - >70% who initially agreed with a blanket consent became unwilling when presented with scenarios of controversial research uses (De Vries et al., 2016)

- Donors reported broad consent provided the right amount of information regarding their decision to donate biospecimens (Warner et al., 2018)

Informed consent: what to disclose

- What to disclose re: data sharing (Institute of Medicine. 2015)
  - Under what conditions the trial data may be shared beyond the trial team;
  - The potential risks to privacy associated data sharing and the protections employed to mitigate this risk
    - 'not legalese weasel-words about "trying hard" to maintain anonymity' (Church in P3G Consortium et al., 2009)

Open or controlled access

- Open access may increase both scientific progress and risks
  - Limited available data to quantify the relative benefits or risks

- Data use agreements
  - significant normative, symbolic, and deterrent value, setting professional expectations and standards for responsible behavior
  - Questions about enforcement (Institute of Medicine 2015)

"Access to individual participant data and trial documents should ***be as open as possible and as closed as necessary****, to protect participant privacy and reduce the risk of data misuse*." (Ohmann et al., 2017)

Open or controlled access

- Open access may increase both scientific progress and risks
  - Limited available data to quantify the relative benefits or risks

"[Open access] is appropriate and desirable for clinical trial results, and … may be the preferred approach when all stakeholders involved in a trial (i.e., sponsors, investigators, and participants) are comfortable with this approach and believe the benefits outweigh the risks.

In many cases, however, sponsors, investigators, and/or participants may have concerns about an open access model for certain clinical trial data and may wish to place some conditions on access to or uses of the data"
(Institute of Medicine 2015)

45CFR 46 2018 Requirements

§46.116 (b)(9) One of the following statements about any research that involves the collection of identifiable private information or identifiable biospecimens:
(i) A statement that **identifiers might be removed** from the identifiable private information or identifiable biospecimens and that, after such removal, the **information or biospecimens could be used for future research studies or distributed to another investigator** for future research studies without additional informed consent from the subject or the legally authorized representative, if this might be a possibility; or
(ii) A statement that the subject's information or biospecimens collected as part of the research, even if identifiers are removed, **will not be used or distributed for future research studies**.

§46.116 (c)(7) A statement that the subject's biospecimens (even if identifiers are removed) may be **used for commercial profit** and whether the subject will or will not share in this commercial profit;

§46.116 (d) Elements of **broad consent for** the storage, maintenance, and secondary research use of **identifiable private information** or identifiable biospecimens. Broad consent for the storage, maintenance, and secondary research use of identifiable private information or identifiable biospecimens (collected for either research studies other than the proposed research or nonresearch purposes) is permitted as an alternative to the informed consent requirements in paragraphs (b) and (c) of this section.

Other regulations

- Other types of regulations might apply, e.g.,
  - HIPAA
  - GINA
- Consideration of needs for new types of protections

NIH data sharing policy

Applications for Receipt Dates BEFORE Jan 25 2023

Applications for Receipt Dates ON/AFTER Jan 25 2023

NIH has issued the Data Management and Sharing (DMS) policy (effective January 25, 2023) to promote the sharing of scientific data. Sharing scientific data accelerates biomedical research discovery, in part, by enabling validation of research results, providing accessibility to high-value datasets, and promoting data reuse for future research studies.

Under the DMS policy, NIH expects that investigators and institutions:

- Plan and budget for the managing and sharing of data
- Submit a DMS plan for review when applying for funding
- Comply with the approved DMS plan

Individual NIH Institutes, Centers, or Offices may have additional policies and expectations (see NIH Institute and Center Data Sharing Policies).

Supplemental Information to the NIH Policy for Data Management and Sharing: Protecting Privacy When Sharing Human Research Participant Data

Notice Number:

NOT-OD-22-213

NIH data sharing policy

Applications for Receipt Dat
BEFORE Jan 25 2023

NIH has issued the Data Manager
data. Sharing scientific data accele
accessibility to high-value dataset

Under the DMS policy, NIH expect

- Plan and budget for the mana
- Submit a DMS plan for review
- Comply with the approved DI

Individual NIH Institutes, Centers,
Sharing Policies).

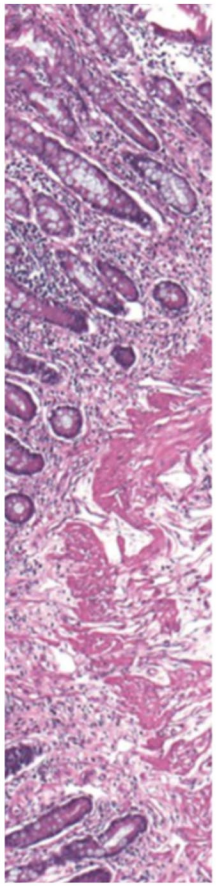
omote the sharing of scientific
of research results, providing

nstitute and Center Data

Supplemental Information to t
Sharing Human Research Pa

Notice Number:
NOT-OD-22-213

aring: Protecting Privacy When



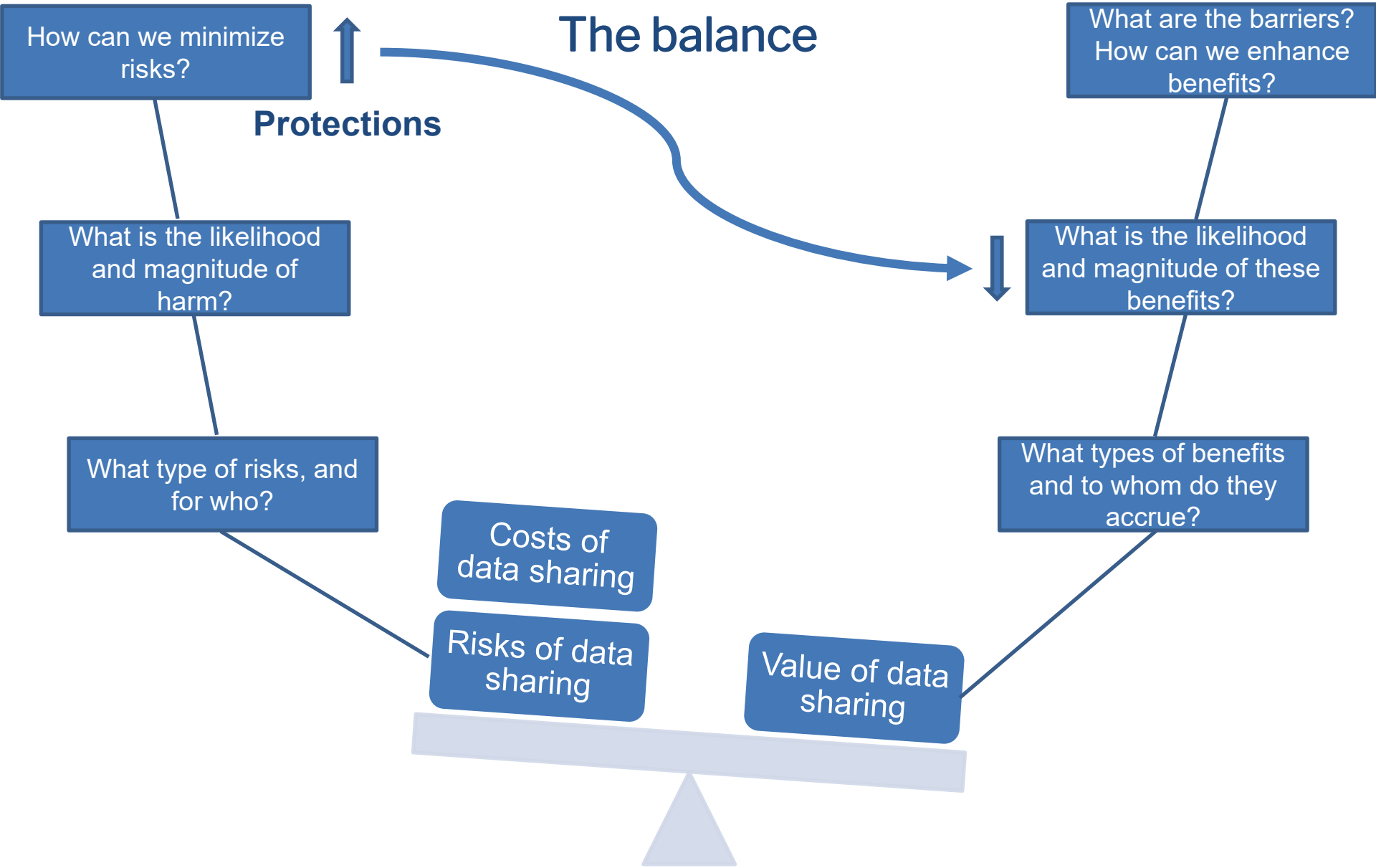
NIH OFFICE OF SCIENCE POLICY
OFFICE OF EXTRAMURAL RESEARCH

INFORMED CONSENT FOR SECONDARY RESEARCH WITH DATA AND BIOSPECIMENS

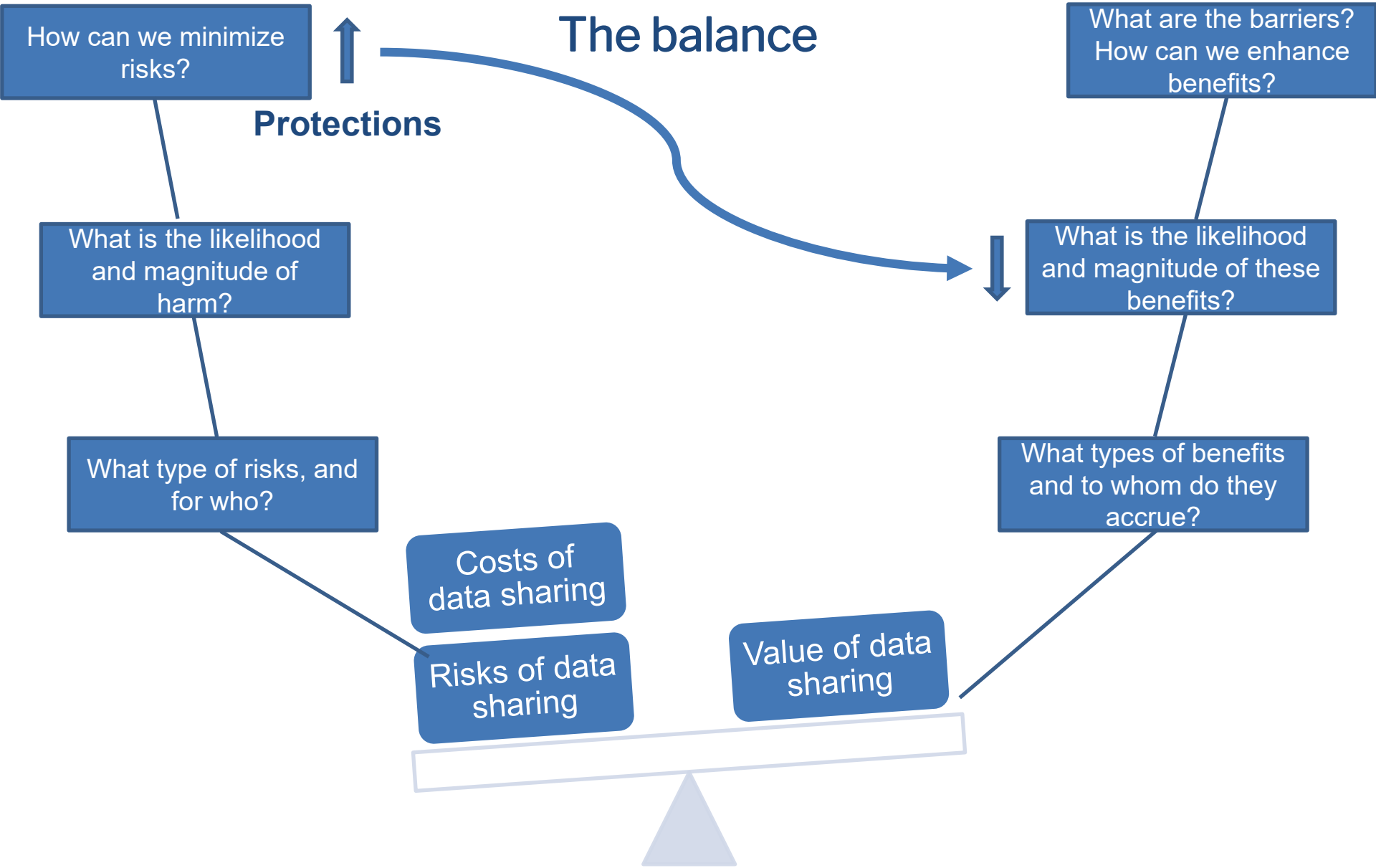
Points to Consider and Sample Language for Future Use and/or Sharing

MAY 2022

The balance


How can we minimize risks?
Protections
What are the barriers? How can we enhance benefits?
What is the likelihood and magnitude of harm?
What is the likelihood and magnitude of these benefits?
What type of risks, and for who?
What types of benefits and to whom do they accrue?
Costs of data sharing
Risks of data sharing
Value of data sharing

The balance


How can we minimize risks?
Protections
What are the barriers? How can we enhance benefits?
What is the likelihood and magnitude of harm?
What is the likelihood and magnitude of these benefits?
What type of risks, and for who?
What types of benefits and to whom do they accrue?
Costs of data sharing
Risks of data sharing
Value of data sharing

Proportionality

Take home messages

"*Data sharing is clearly, in general, a good idea. However, it is not as simple as it seems. The devil is in the detail, and the detail is highly specific to each study, and each potential data recipient.*"
(Pearce & Smith, 2011)

Thank you!

Saskia.Hendriks@nih.gov