Data Sharing Dilemmas

Michelle Mello, JD, PhD

Stanford MEDICINE | Department of Medicine
Center for Health Policy/
Primary Care and Outcomes Research

Stanford Law School

Disclosures

- Advisor to Verily Life Sciences, LLC on the design of a product facilitating safe return to work and school during COVID-19
- Collaborator on studies using MPOG data

Potential benefits of data & code sharing

- Facilitate secondary analyses to accelerate science
- Heighten the ability to evaluate and reproduce studies
- Strengthen the culture of sharing and openness in science
- Conserve scarce research funding
- Increase opportunities for teaching and learning
- Honor patients' and research participants' contributions

Data sharing dilemmas

Meaningful access

Burdens and incentives

Patient privacy and consent

Accountability

Data sharing policies

Funders in 2019:

- 38% of noncommercial funders had a DS policy; 60% of those *required* DS
- 41% of commercial funders had a policy committing to share their own data
- Policies only apply to *clinical trials*

Source: Gaba et al., PLoS ONE 2020.

Surgery journals in 2019:

- Among 82 journals publishing trials, 5% required DS *statement* and 45% encouraged one
- Statement need not express intent to share IPD
- Policies only apply to *clinical trials*

Source: Lombard N et al., Trials 2020.

How is “access” commonly interpreted?

Original Investigation | Health Policy

January 28, 2021

Evaluation of Data Sharing After Implementation of the International Committee of Medical Journal Editors Data Sharing Statement Requirement

Valentin Danchev, DPhil[1,2,3]; Yan Min, MD[4]; John Borghi, PhD[5]; et al

» Author Affiliations | Article Information

JAMA Netw Open. 2021;4(1):e2033972. doi:10.1001/jamanetworkopen.2020.33972

Examined 487 clinical trial reports in *JAMA*, *Lancet*, & *NEJM* 2018-2020

Results A total of 334 of 487 articles (68.6%; 95% CI, 64%-73%) declared data sharing, with nonindustry NIH-funded trials exhibiting the highest rates of declared data sharing (89%; 95% CI, 80%-98%) and industry-funded trials the lowest (61%; 95% CI, 54%-68%). However, only 2 IPD sets (0.6%; 95% CI, 0.0%-1.5%) were actually deidentified and publicly available as of April 10, 2020. The remaining were supposedly accessible via request to authors (143 of 334 articles [42.8%]), repository (89 of 334 articles [26.6%]), and company (78 of 334 articles [23.4%]). Among the 89 articles declaring that IPD would be stored in repositories, only 17 (19.1%) deposited data, mostly because of embargo and regulatory approval. Embargo was set in 47.3% of data-sharing articles (158 of 334), and in half of them the period exceeded 1 year or was unspecified.

Dilemmas

Meaningful access

Burdens and incentives

Patient privacy and consent

Accountability

Potential commercial harms

- Competitive harm: could help competitors develop competing products or gain a marketing advantage
- Use in regulatory enforcement or litigation
- Patent issues:
  - Early disclosure may be “prior art” that could jeopardize patentability
  - May also trigger the patent period to begin running
  - Arms competitors with information to beat the discloser to be first to file
- Cost burdens:
  - Statistical and clinical staff to evaluate data requests
  - Legal staff to draft and enforce DUAs
  - Technical staff to prepare data

Who uses shared data?

- **dbGaP**: 82% not-for-profits
- **European Medicines Agency, 2010-2012**: 39% industry or consultant,16% lawyers, 8% academics
- **YODA (J&J data)**: 90% academic, 5% industry

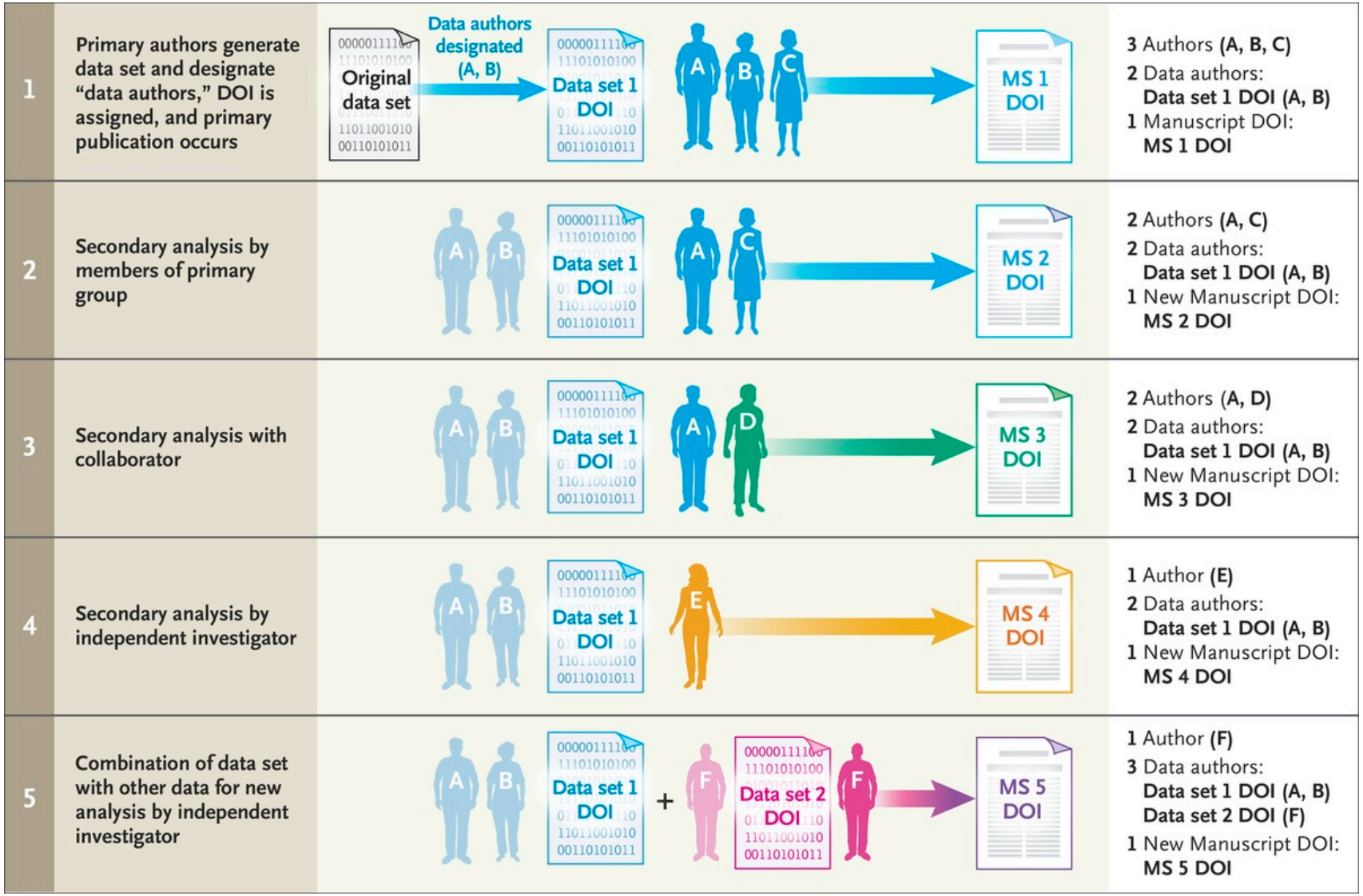
Other incentives problems

- Fairness concerns
- Reduced incentives to contribute data
- Potential to be “scooped” on subsequent analyses
- The PITA factor
- Acknowledgment vs. coauthorship

Potential incentives

- User fees
- Institutional recognition
- Individual authorship credit

Data authorship



Source: Bierer BE et al. N Engl J Med 2017;376:1684-1687.

An individual researcher (letters A through F) may be designated and credited as an **author**, a **data author**, or both, depending on the person's contribution to the data and analysis in the published work.

DOI = digital object identifier MS = manuscript

Dilemmas

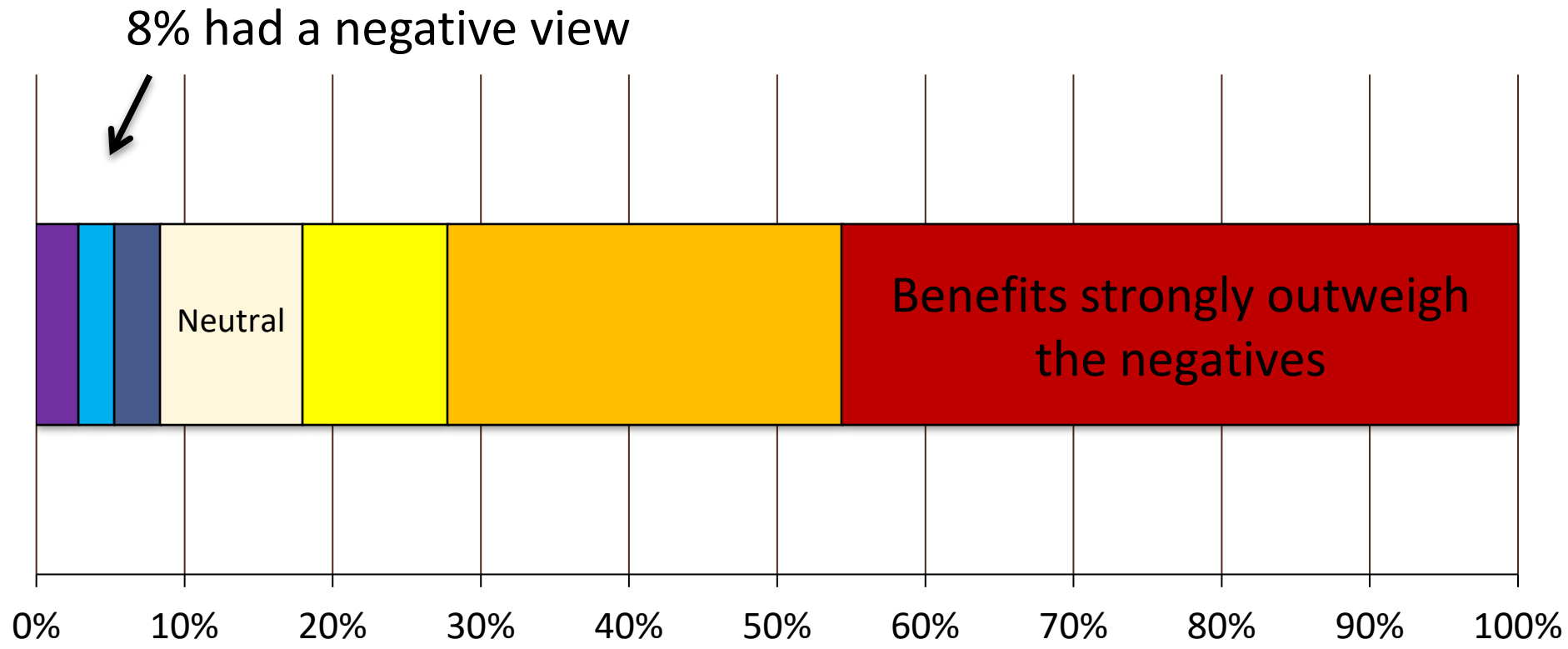
Meaningful access

Burdens and incentives

Patient privacy and consent

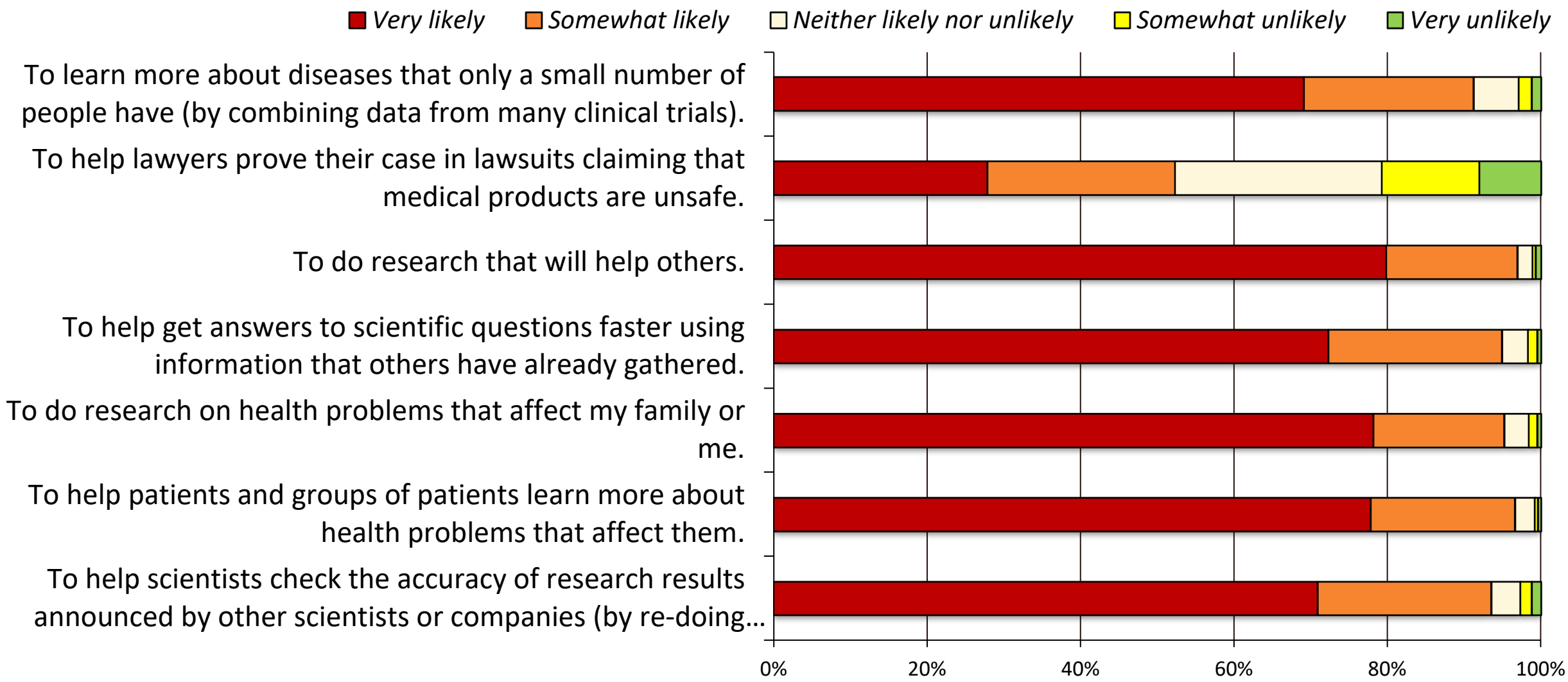
Accountability

Overall, how do you think the potential benefits of sharing anonymous, individual clinical trial data weigh against the potential negative consequences?



Source: Mello MM et al., N Engl J Med 2018

How likely would you be to allow your anonymous, individual clinical trial data to be used in the following ways?



Source: Mello MM et al., N Engl J Med 2018

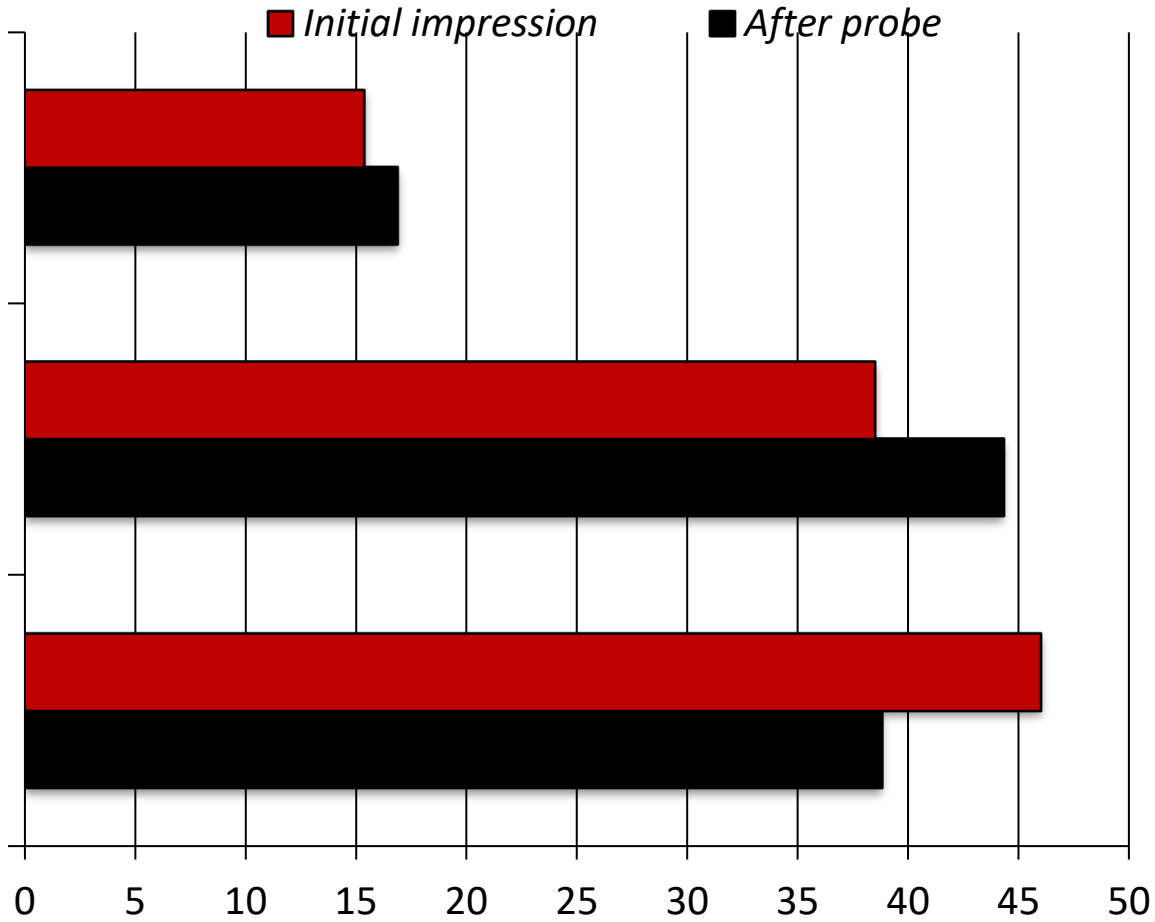
Which of the following best describes how you would feel about being asked for permission to share your anonymous, individual clinical trial data with people outside of the trial?

PROBE: Suppose it's true that *when more people decide not to share their individual clinical trial data, the scientific value of the remaining data is lower.* Please think about whether that would change your answer.

As long as the researchers have good data security protections, data sharing doesn't need to be discussed in the informed consent form.

The informed consent form should explain that my data are going to be shared (and tell me how the researchers will protect my identity).

The informed consent form should ask me for specific permission to share my data, separate from my overall consent to be in the clinical trial (and tell me how the researchers will protect my identity).



Source: Mello MM et al., N Engl J Med 2018

Table 3. Notification and Permission Preferences for Research on Medical Practices

Response	Research Scenario (n = 1095), n (%)		
	Medical Record Review	Randomization (Hypertension)	Randomization (Serious Condition)
"If you were newly diagnosed with high blood pressure and this research were happening in your health system, how would you prefer to be notified about this research?"			
No notification	109 (10.0)		
General information	162 (14.8)		
Discussion plus verbal permission	266 (24.2)		
Discussion plus written permission	558 (51.0)		

27% of this group would prefer research not be conducted if getting permission isn't feasible

Source: Cho MK et al., Annals Intern Med 2015

Dilemmas

Meaningful access

Burdens and incentives

Patient privacy and consent

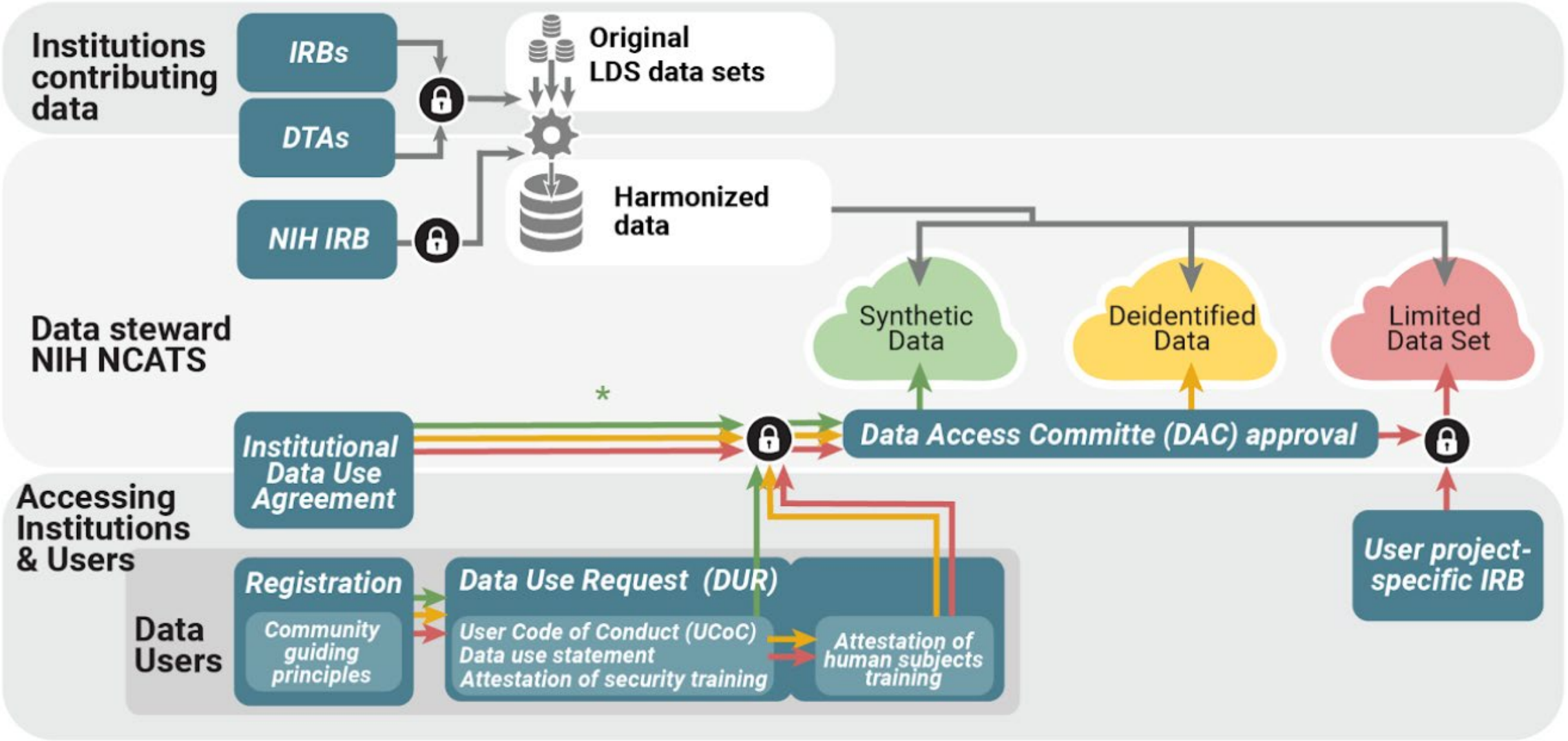
Accountability

Accountability strategies

- Strong DUAs
- Incorporate patients into decisions about data uses and transfers
- “Minimum necessary” principle
- Use of “honest brokers”

N3C Model

Source: Emery IF, Presentation 1/25/21.



N3C Data Levels

Data Level	Data Description	Eligible Users	Access Requirements*
Synthetic Data Set	Artificial, statistically-comparable, computational derivative of the original data: it does not contain individually identifiable health information, also known as protected health information (PHI) as defined by the Health Insurance Portability and Accountability Act (HIPAA)	• Researchers from U.S.-based institutions • Researchers from foreign institutions • Citizen scientists	• N3C registration • N3C Data Enclave account • Data Use Agreement (DUA) executed with NCATS • NIH IT training completion • Approved Data Use Request (DUR)
De-identified Data Set	Patient data that has been stripped of PHI identifiers as defined by HIPAA	• Researchers from U.S.-based institutions • Researchers from foreign institutions	• N3C registration • N3C Data Enclave account • DUA executed with NCATS • NIH IT training completion • Approved DUR • Human Subjects Research Protection training completion
Limited Data Set	Patient data that includes only two of the 18 elements defined as PHI by HIPAA (dates of service and patient zip code)	• Researchers from U.S.-based institutions	• N3C registration • N3C Data Enclave account • DUA executed with NCATS • NIH IT training completion • Approved DUR • Human Subjects Research Protection training completion • Local Human Research Protection Program IRB determination letter

Source: ncats.nih.gov/n3c

Conclusions

Current requirements are minimal and loosely enforced

While security issues are real, patients support sharing IPD.

For EHR data, need to find ways to honor the desire to give consent.

Incentives problems are systemic, but MPOG could adopt expansive attribution policies.

Promising systems exist, though technological strategies require resources.