# Data Sharing in Research: Challenges and Opportunities
# Are we Ready Yet (for the risks)?

Robert M Califf MD

Vice Chancellor for Clinical and Translational Medicine

Duke University

## Key Points

- We need to be ready to share research data transparently
- All you have to do is say: “I’m in favor of secret human experimentation” to know why.
- But there are many issues to work through before the system could be turned on
  - Researcher issues
  - Business issues
  - Regulatory issues
  - Patient issues

## The Premise

- Trial participants give consent to be involved in a human experiment
- The basis for the experiment that enables ethics committees to approve it is the commitment to create generalizable knowledge
- The results should be reproducible by independent parties
- First base: publishing the results
- Second base: sharing summary data
- Third base: confidentially sharing detailed individual participant clinical data
- Home run?: publicly sharing detailed clinical data

The NEW ENGLAND JOURNAL of MEDICINE

HEALTH LAW, ETHICS, AND HUMAN RIGHTS

# Preparing for Responsible Sharing of Clinical-Trial Data

Michelle M. Mello, J.D., Ph.D., Jeffrey K. Francer, J.D., M.P.P., Marc Wilenzick, J.D., Patricia Teden, M.B.A., Barbara E. Bierer, M.D., and Mark Barnes, J.D., LL.M.

HARVARD | SCHOOL OF PUBLIC HEALTH

# Core Principles

1. Provide access sufficiently broad to achieve the sought-after benefits
   - At a minimum, prospectively apply to trials of all drugs, devices, and biologics approved in at least one country
2. Ensure responsible use
3. Protect participants' privacy
4. Treat all qualified data requesters and trial sponsors evenhandedly

# Core Principles, continued

5. Hold data requesters and generators accountable
   - Data requesters should precommit to an analytic plan
   - Transparent, principled decisions about data releases

6. Be practicable
   - Timely decisions
   - Avoid undue burdens on data generators
   - Able to handle large volume of trials
   - Globally harmonized standards

# 8 Pillars of a Sound Data Sharing System

1. A binding mechanism to ensure universal participation
2. Minimum standards for what must be shared, and how
3. Any requirements apply equally to all trial sponsors
4. Explicit decision criteria for data releases
5. Public disclosure of reasons for decisions
6. Public disclosure of requesters' identities and plans
7. Mechanism to enforce conditions of data use
8. Technical support

# Four Potential Models

| Open Access Model | |
|---|---|
| **Decision-maker** | None. |
| **Criteria** | **Responsible-use attestation**: All requests granted if Requester attests that data will not be used inappropriately (e.g., to reidentify research participants) |
| **Process** | • Data Generator routinely posts data from trials when results are publicly reported or submitted to regulator, along with documentation to facilitate use of data<br>• Researchers can simply download the material |

# "Black Box" / Database Query Model

| | |
|---|---|
| **Decision-maker** | Independent review board or Data Generator |
| **Criteria** | 1. Sound science<br>2. Benefit/risk balancing<br>3. Expertise |
| **Process** | • Requester submits a research query to the Data Holder<br>• Data Holder runs the query and returns results—not data |

# Data Generator Model

| | |
|---|---|
| **Decision-maker** | Data Generator |
| **Criteria** | 1. **Sound science**: Is there a reasonable scientific hypothesis, sound analytical plan, and adequate plan to disseminate findings?<br>2. **Benefit/risk balancing**: Do the potential public health benefits of answering the proposed question(s) outweigh the probable adverse effects on the Data Generator (intellectual-property interests, competitive concerns, technical-support burden) and the potential risks to research participants?<br>3. **Expertise**: Does the research team have expertise sufficient to carry out the proposed analyses? |
| **Process** | • Data Generator reviews request, decides, and publicly documents rationale for decision.<br>• Denials are appealable to independent Appellate Board, whose decision is final. |

| | Data Generator Model | Learned Intermediary Model |
|---|---|---|
| **Decision-maker** | Data Generator | Review Board that is independent of Data Generator |
| **Criteria** | 1. **Sound science**: Is there a reasonable scientific hypothesis, sound analytical plan, and adequate plan to disseminate findings?<br>2. **Benefit/risk balancing**: Do the potential public health benefits of answering the proposed question(s) outweigh the probable adverse effects on the Data Generator (intellectual-property interests, competitive concerns, technical-support burden) and the potential risks to research participants?<br>3. **Expertise**: Does the research team have expertise sufficient to carry out the proposed analyses? | |
| **Process** | • Data Generator reviews request, decides, and publicly documents rationale for decision.<br>• Denials are appealable to independent Appellate Board, whose decision is final. | • Board reviews request, collects input from Data Generator, decides, and publicly documents rationale for decision |

# Three Questions You Should Not Address on an Empty Stomach

- How much should companies' **(or health systems)** commercial concerns be credited?
- Can a system in which Data Generators control release decisions be trustworthy—and trusted?
- Is a regulatory mandate desirable?

# Food for Thought

- Data sharing isn't free. Who should pay?
- Can data generators' concerns be addressed through restrictions on *use* of data, rather than restrictions on *access*?
- What obligations do regulators assume to act in response to new analyses—and what's needed to enable them to do it?

## Problems Now

- Researchers
  - Important trials, not so important trials
  - Rewards for the work
- Business and government agencies
  - Cost of preparing the data sets
  - Risk of rogue analyses
- Regulators
  - Regulatory dysfunction
  - Global disharmony
- Patients/public
  - Privacy

# All Trials, Really?

- Over the 300 new trials per week in ClinicalTrials.gov
- Most are small and not designed to answer a clinically important question
  - Vast majority have < 100 participants
- Doesn't include growing number of trials done completely outside of US
- How can you decide ahead of time which trials are important enough to be included?

## Rewards for Work

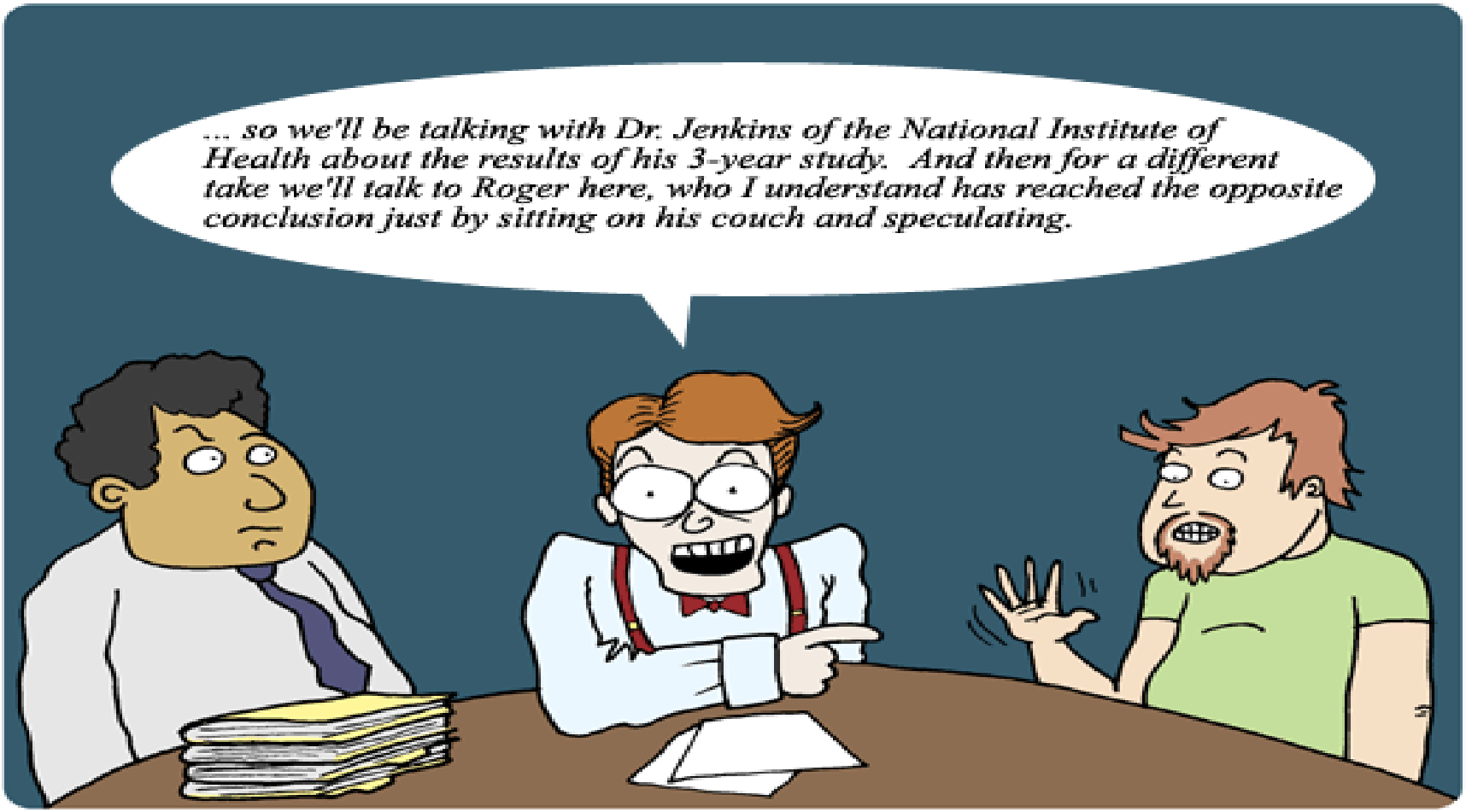
- Doing clinical trials is hard work
- Having your work taken is frustrating enough when it is done well
- We need more people involved in the primary work of generating reliable evidence

## Cost

- Preparing and maintaining a data set is expensive
- Currently, publicly funded trials are already underfunded
- Results reporting in ClinicalTrials.gov, which is required is not so good, especially in public sector

## Rogue Analyses

- It is not hard to take segments of data and produce sheer random snapshots that look compelling out of context (last month's Economist)
  - If 100 people are trolling the internet for study data and doing random analyses, and only the "significant" results are reported, most will be false positive
- Clinical data are very context dependent—it takes time to learn the context and much of it is not written down




http://www.skepticalscience.com/SkS-Weekly-Digest_20.html

# Bottom Line

- I favor setting a goal of 5-10 years and doing careful phased experiments
- First, I would shore up ClinicalTrials.gov to optimize results reporting in that system
  - Especially AMCs and Federal agencies (NIH, VA, etc.)
- Pick some areas and topics for demonstration project funding
  - Different methods of dealing with individual patient data
  - Approaches to industry consortia
  - Approaches to global harmonization
- In absence of careful, step-wise implementation studies I very much favor the learned intermediary
- Use Collaboratory and PCORnet with its DRNs and PPRNs as test cases for the new learning health system world
- Don’t rush into Obamacare website problem!