A CULTURE OF OPEN INQUIRY: PRODUCING CREDIBLE EVIDENCE ON PROGRAM EFFECTS

E. Michael Foster
School of Public Health
UNC-CH

These comments are my own, but I would like to thank Len Bickman.

Premise

We have relatively little meaningful evidence about intervention effectiveness (and cost-effectiveness).

Much (?) of the evidence is of poor quality, largely because of poor methodology, and evidence on failures is buried.

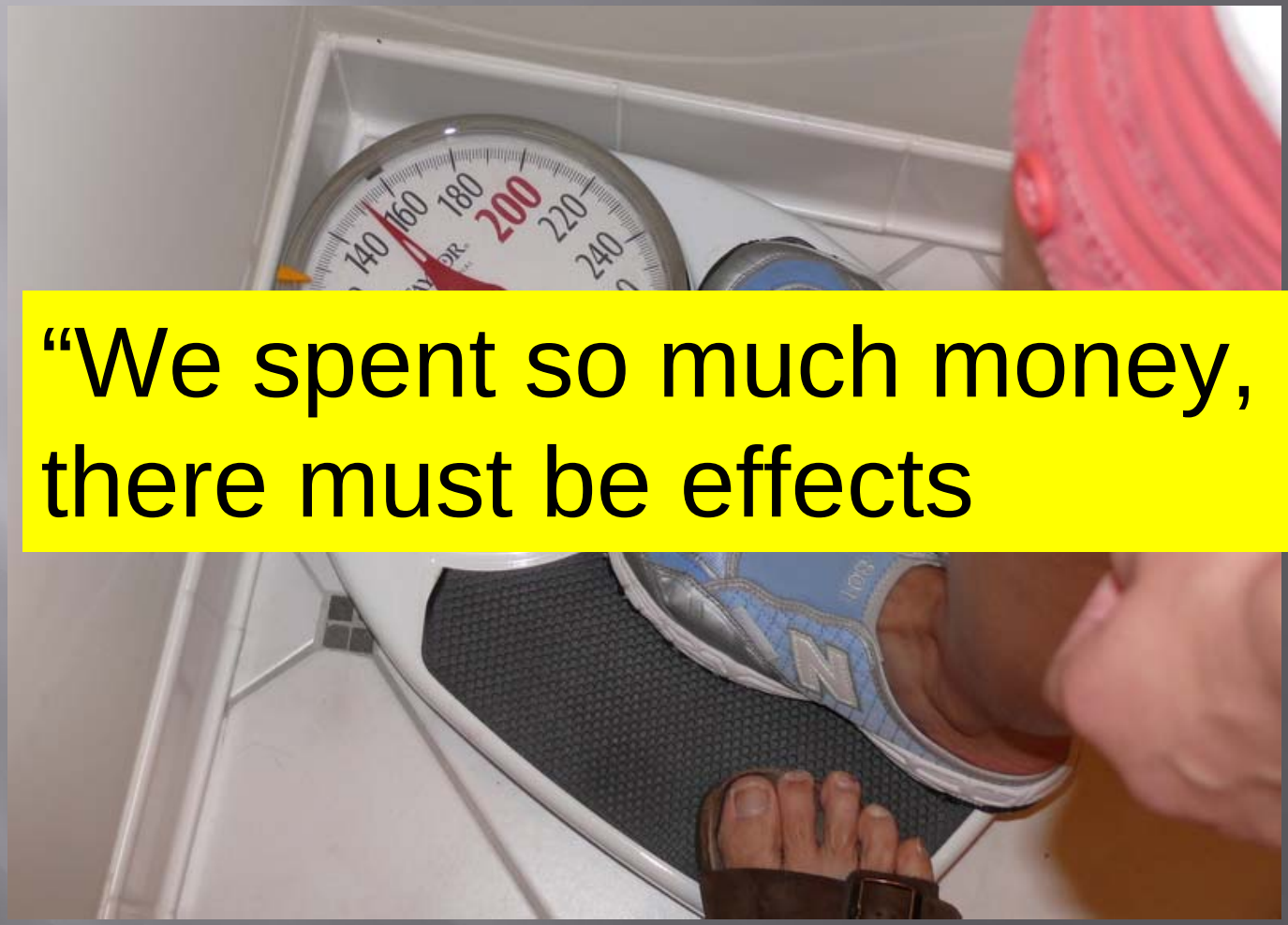
As we move to "scale", A Crisis of Sorts

- **These problems are not going to get a bit better as we move to effectiveness trials**
- **The problems will get worse**
- Examples:
  - Dosage
  - Condition switching (WLC)
  - Fidelity not randomly assigned
  - Differential impact
  - Craziness with administrative data

Outline

- The FDA would not consider much of our research as credible evidence of efficacy
- Outsiders often judge published evidence as poor (case study: MST)
- The published literature is not representative of overall findings
- Very few of these findings would replicate
- The solution: A Culture of Open Inquiry

Every time I turn around…

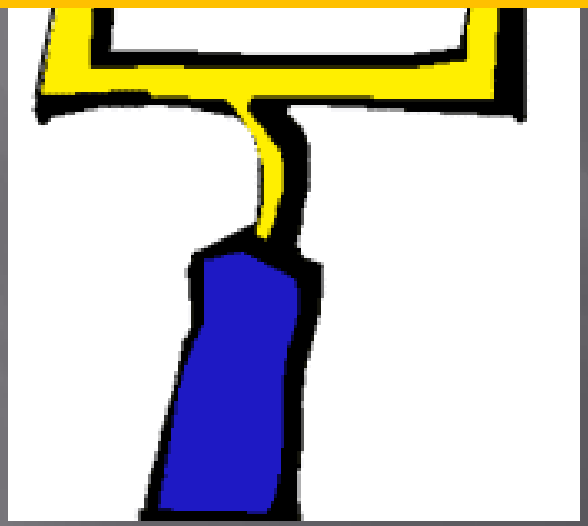


"We spent so much money, there must be effects

Someone has their darned foot on the scale!

"We get effects if we include an interaction with time-squared and we leave out the main effect of that term."

When I talk to other methodologists, they have burst out laughing, saving "I've been told that, too!"



The FDA Would Never Accept Much of the Research as Evidence

FDA guidelines indicate that

"For each clinical trial contributing to a marketing application, all important details of its design and conduct and the principal features of its proposed statistical analysis should be clearly specified in a protocol written **before the trial begins**."

"The extent to which the procedures in the protocol are followed and the primary analysis is planned a priori will contribute to the degree of confidence in the final results and conclusions of the trial."

A Perpetual State of Pilot Studies

"However, in contrast to confirmatory trials, [the] objectives [of exploratory trials] may not always lead to simple tests of predefined hypotheses. In addition, exploratory trials may sometimes require a more flexible approach to design so that changes can be made in response to accumulating results. **Such trials cannot be the basis of the formal proof of efficacy**, although they may contribute to the total body of relevant evidence. "

"Only results from analyses envisaged in the protocol (including amendments) can be regarded as confirmatory."

Assessment of MST

Methodological quality

- Sloppy or least a lack of detail
- Technical problems (e.g., hazard modeling and length of followup period)
- Nearly all studies by developers

Assessment of MST

"There is inconclusive evidence of the effectiveness of MST compared with other interventions with youth. There is no evidence that MST has harmful effects."

Littell JH, Popa M, Forsythe B. Multisystemic Therapy for social, emotional, and behavioral problems in youth aged 10-17. *Cochrane Database of Systematic Reviews 2005, Issue 4. Art. No.: CD004797. DOI: 10.1002/14651858.CD004797.pub4.*

"Why Most Findings are False"

- The smaller the studies conducted in a scientific field, the less likely the research findings are to be true.
- The smaller the effect sizes in a scientific field, the less likely the research findings are to be true.
- The greater the number and the lesser the selection of tested relationships in a scientific field, the less likely the research findings are to be true.

Why Most Findings …

- The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true.
- The greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to be true.

Why Most Findings…

- The hotter a scientific field (with more scientific teams involved), the less likely the research

Ic

Much of prevention research meets all of these criteria. These criteria raise the possibility that my experiences reflect widespread problems.

Fixing the Problem: A Culture of Open Inquiry

In general, what we need is an open process, informed by relevant technical expertise.

I) Data Sharing: Why?

- Provides essential detail
- Provides opportunity to examine robustness and representativeness
- Establishes accountability

1) Data Sharing: How?

- Protect confidentiality
- Widespread availability (not three collaborators)
- Three levels
  - Data used in published analyses
  - Data used to execute analysis plan in grant
  - All of the taxpayer's data

II) A *Sound* Analysis Plan

- Match the design of the study

(as implemented, not hoped for)

- Adjusts for multiple testing
- Includes cross-validation
  - When model specification is data driven
  - When unexpected findings appear
- Deals seriously with the possibility of failure at the start of the study
- *Good data quality* (often "amateurish")

III. Proper Reporting

- The basic analysis in the analysis plan should be reported fairly quickly.
  - Determined the power
  - Clearly set before one examines the data
- If you tried it both ways, report both ways.
- May include multiple outcomes but all are reported

IV) More methodology within prevention

The good news is that there are solutions available for dealing with many challenging problems, but

- Quantitative training in psychology is inferior to that in other social sciences* (I can document this)
- More biostatisticians and social scientists from other fields are needed

V. Better Reviewing

Be very skeptical if you are reviewing a paper from a project that reports

- results from five years earlier
- transformed variables but not the ones in the original scale.
- terribly nuanced findings

Ask yourself: Is nuance informative?

VI. Broader Mix of Investigators

- Skeptics needed; True believers, not so much.
- Researchers from the same background share the same weaknesses
- *The skills that make one good at developing interventions are very likely not the ones that are good for evaluating them.*
- Senior methodologist collaborators

Good Examples

Data sharing

- Incredible Years
- Abcederian
- New Hope

Mix of Investigators

- *MTA*

Other Aspects

- Accountability by NIH—real reports on what's in the original analysis plan or the $$$ stops
- Clinical trials data center
- More money for analyses and analysts
- Journals can really help by
  - Publishing null findings
  - Requiring data sharing

JHR Policy on Replication and Data Availability

1. Manuscripts submitted to the JHR will be judged in part by whether they have reconciled their results with already published research on the same topic. In cases where a past study has obtained different results for reasons that are not obvious on an a priori basis, authors may be required to perform some comparative estimation with their own data set. In addition, the JHR will continue its existing policy of requiring authors to present the results of sensitivity tests.

2. Authors of accepted manuscripts will be asked to preserve the data used in their analysis and to make the data available to others at reasonable cost from a date six months after the JHR publication data and for a period of three years thereafter. Authors wishing to request an exemption from this requirement should notify the editors at the time

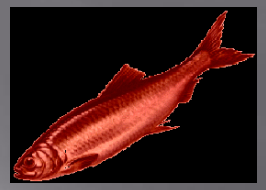
Barriers

- Self interest—hoarding data is indeed good for one's career
- Self interest—testing new interventions is more rewarded than evaluating existing ones
- Technical expertise (e.g., data quality and analysis)
- Scientific interest

(ex: sample representativeness)

Barriers

- NIH would need more resources for monitoring findings and progress
- Data privacy (HIPAA)



Objections

- "I put my heart and soul into collecting data"
- "If I start sharing my data, I'm going to have to document the data"
- "What? Should I just let other people write half-ass papers with my data?"
- "Other people don't understand my product well enough to analyze the data"
- "I'm doing what I'm supposed to do"

Objections

- "You can have my data but I ain't giving you any money" or "I'd love to have more biostat input—but have you priced one lately!!!"
- "How in the world am I supposed to capture the complexity of what we're doing in advance in an analysis plan"
- "What we do is just more complicated than drug trials. You just don't get it."