



Rethinking the Rush to Randomize: A Brief History of RCTs in Government and Politics, Implications, and Concerns

Michael F. Crowley
Consultant
(Former Senior Policy Analyst,
White House Office of Management and Budget)



Rethinking the Rush to Randomize:

Summary

- Evaluating for program effectiveness under the Bush Administration
- Obama Administration developments
- Intersection with Politics
- Concerns and Risks
- Need for a New Evidence-Based Focus or Paradigm

Rethinking the Rush to Randomize:

What is OMB?

- OMB is a White House agency responsible for:
 - ✓ Developing the President's Budget
 - ✓ Overseeing agency management improvements
 - ✓ Making program and policy recommendations
- OMB has long played a role in assessing program performance and effectiveness:
 - ✓ Traditional spring reviews of selected agency programs
 - ✓ Periodic formal and ad hoc program reviews
 - ✓ Issuance of guidance on performance measurement, program assessment, and evaluation



Rethinking the Rush to Randomize

A Bit of History



Rethinking the Rush to Randomize: Bush Administration

- Program Assessment and Rating Tool (PART) launched in 2002
- PART assumed “we do not know what we are getting for our money”
- PART used to evaluate programs and issue ratings of program effectiveness.
- Five ratings: effective, moderately effective, adequate, ineffective, and “results not demonstrated”

Rethinking the Rush to Randomize: Bush Administration (cont'd)

- Education develops “what works” clearinghouse in 2002 and issues NCLB regs in 2003 – focus on scientifically-based evaluation
- OMB revises the PART in 2004 to include questions on “rigorous evaluation”
- OMB issues guidance in 2004, “What Constitutes Strong Evidence of a Program’s Effectiveness?” – emphasis on RCTs
- With agency, GAO, and AEA pushback, OMB tables plans to issue additional guidance

Rethinking the Rush to Randomize: Obama Administration

- OMB discontinued the PART
- OMB issues Memo – *Increased Emphasis on Program Evaluations* (M-10-01), October 7, 2009
 - ✓ Dedicated funding for evidence-based priorities, enhanced agency evaluation capacity, new rigorous evaluations
 - ✓ Interagency working group to promote stronger evaluation
- OMB issues – *Evaluating Programs for Efficacy and Cost-Efficiency* (M-10-32), July 29, 2010
 - ✓ Focus on evidence for agency priorities, agency evaluation plans, and assessments of agency evaluation capacity
- OMB issues – *Use of Evidence and Evaluation in the 2014 Budget* (M-12-13), May 18, 2012
 - ✓ Explicit references to random assignment; implicit preferences for RCTs

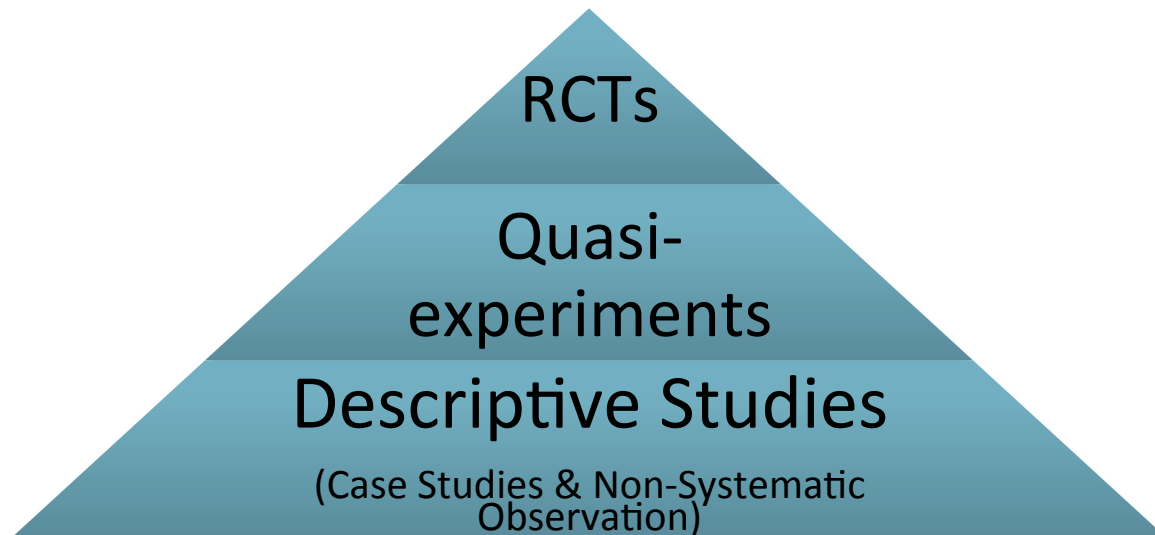
Rethinking the Rush to Randomize: Obama Administration Themes

- Review and use existing evidence of what works
- Conduct rigorous evaluation to build evidence
- Scale up proven, effective practices
- Expand and increase use of innovative tools to build evidence and invest in “what works”:
 - ✓ Expansion of “what works” clearinghouses: Justice (2011) and Labor (2013) and push for alignment
 - ✓ Use of evidence-based multi-tiered grants (e.g., Education’s Investing in Innovation “i3” Fund and CNCS’ Social Innovation Fund) (2009+)
 - ✓ Pay for Success Grants, multiple agencies (2012+)
 - ✓ Performance Partnerships (2012+)
- Implicit preference for RCTs and assumptions of evidence hierarchy are core to the initiatives

Rethinking the Rush to Randomize: Finding Evidence of What Works

- RCTs are preferred
- Notions of evidence hierarchies in which “stronger” evaluation designs trump “weaker” ones

Evidence Hierarchy





Rethinking the Rush to Randomize: Interest in “What Works” is Bipartisan

- Rep. Paul Ryan (R-WI) proposed a federal Commission on Evidence-Based Policy Making (2014)
- Heritage Foundation suggested that Congress should require experimental evaluations for any programs authorized or reauthorized (2011)
- Concern for “what works” has factored into political debate about Head Start (2010+), VAWA (2012+) and Medicaid Expansion under the Affordable Care Act (2013+)



Rethinking the Rush to Randomize

What's Wrong With the Model?

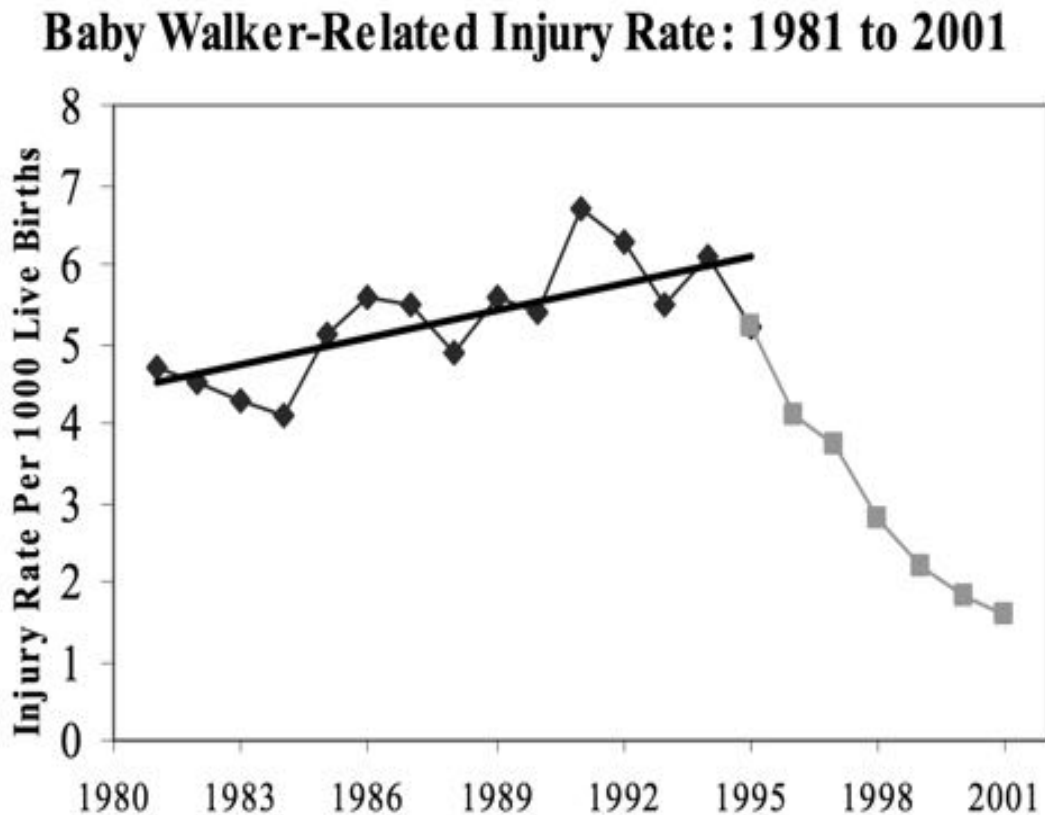


Rethinking the Rush to Randomize: A Problem—Other Methodologies

- RCTs often are **unnecessary** for making informed, analytic determinations of whether programs are effective or not
- Often, **deeper insights** are available from other methodologies

Rethinking the Rush to Randomize: Baby Walkers

- OMB formed an interagency work group to write guidance on when to use RCTs (2004). Project discontinued but produced interesting discussion:



Was the CPSC standard effective in reducing injuries?

Evaluation Methodology: Interrupted time series. No control group, but alternative explanations controlled through data analysis and logic (no decrease in walker sales).

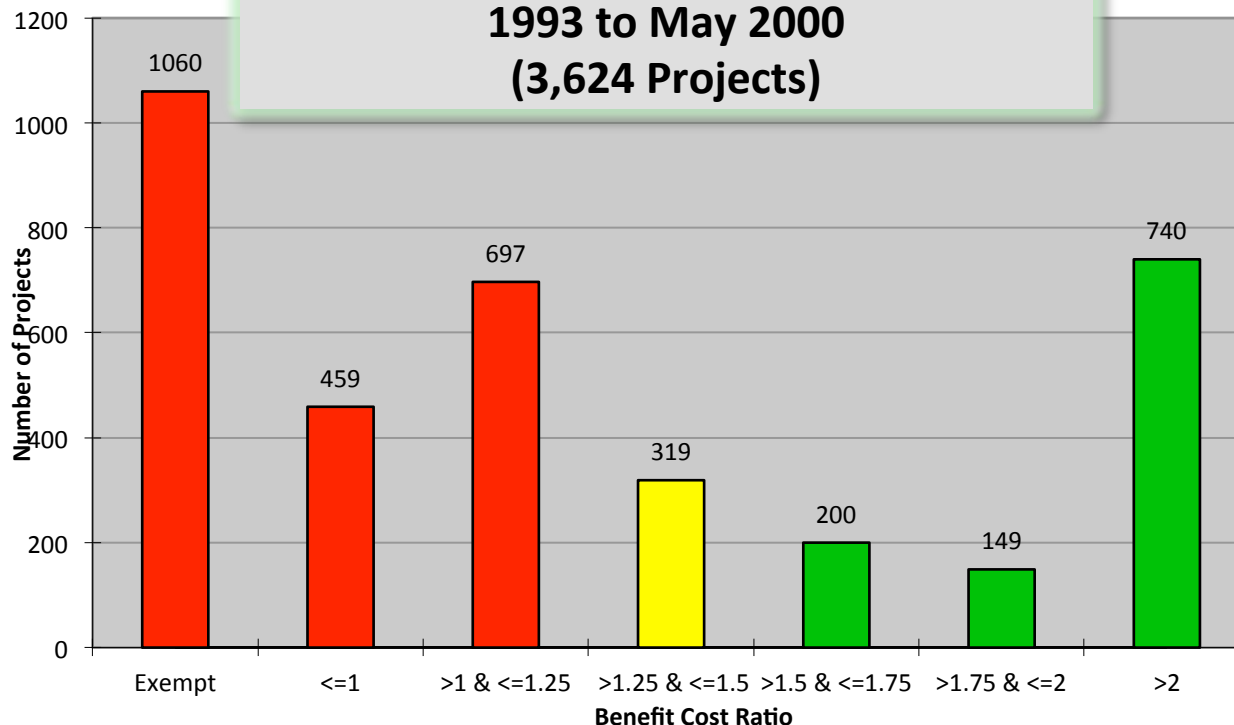
Random assignment not ethical or necessary.

Rethinking the Rush to Randomize:

FEMA Hazard Mitigation Grants

- OMB analyzed benefit cost ratios – 61% of projects were questionable or demonstrably ineffective

**Benefit Cost Ratios for FEMA
Hazard Mitigation Grant Program Projects
1993 to May 2000
(3,624 Projects)**



Were FEMA projects producing more benefit for the taxpayer than cost?

Evaluation Methodology: Analysis of program data supplemented by interviews with program staff and limited field visits. No control group necessary – all non-exempt projects examined for key data.

Random assignment not feasible or necessary.

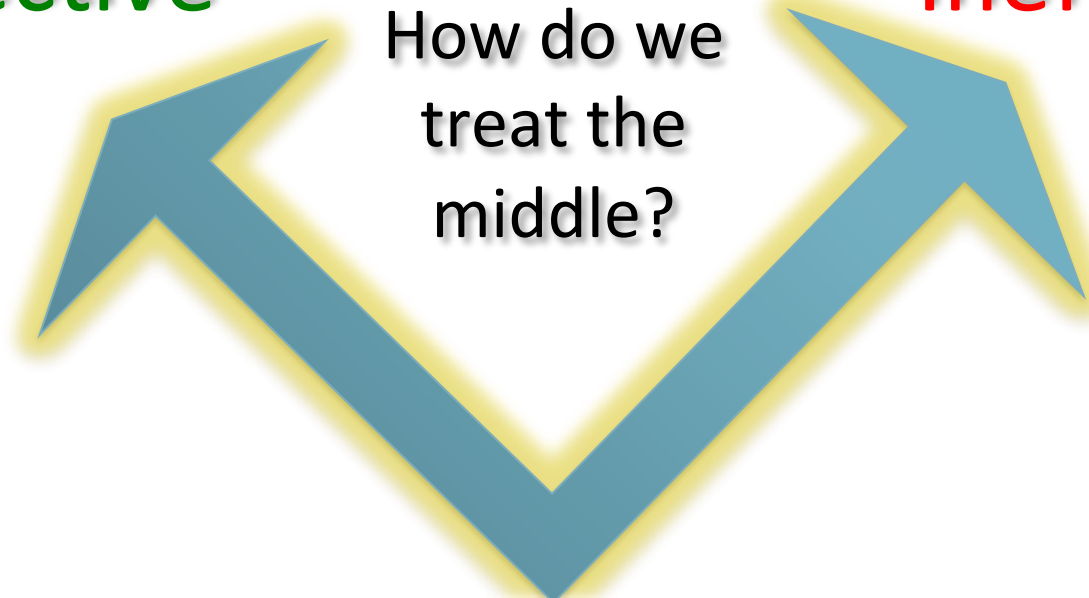
Rethinking the Rush to Randomize: Another Problem—Binary Thinking

- The evidence-based movement in government fosters binary thinking (effective or ineffective, support programs or not, cut program funding or don't cut) and ignores evaluation concerned with **program improvement**

Effective

Ineffective

How do we
treat the
middle?





Rethinking the Rush to Randomize: Another Problem—Risk of Lopsidedness

- Focusing all of our attention on “what works”, government decision-makers and policymakers fail to consider:
 - ✓ How do programs work?
 - ✓ When do (or do not) programs work?
 - ✓ Why (or why not) do programs work?
 - ✓ What are the issues in program design/operation that we should consider from efficiency and equity perspectives?
 - ✓ Are programs cost effective?
- Evaluation that limits itself to questions RCTs are good at probing gives up too much fertile ground for analysis and understanding



Rethinking the Rush to Randomize: Another Problem—Loss of Big Picture

A narrow focus on “what works”, the effectiveness of specific interventions, e.g.:

- ✓ school curricula
- ✓ policing tactics
- ✓ teen pregnancy strategies

Can draw our attention away from larger, entrenched, more fundamental issues, e.g.:

- ✓ inequity of school finance
- ✓ high incarceration rates
- ✓ problems with access to health care

Rethinking the Rush to Randomize Another Problem—Model is Outdated

- The government evidence-based movement is where evidence-based medicine was 20 years ago (stress on RCTs and evidence hierarchies)
- Today, evidence-based medicine is moving away from rigid evidence hierarchies and single-minded focus on single methodologies toward systematic reviews informed by:
 - ✓ multiple research methods
 - ✓ expert opinion
 - ✓ practitioner feedback
 - ✓ judgment



Rethinking the Rush to Randomize: Another Problem—RCTs are difficult

- RCTs only work well with discrete interventions and well defined outcomes (construct validity)
- In the social sciences and government, most:
 - ✓ RCT are difficult to judge because effects sizes are too small
 - ✓ RCTs are underpowered – sample sizes are too small
- Many RCTs suffer from any number of design flaws and biases
- Even if RCTs are good, they need replication
- Even if replicated, social and cultural factors tend to shift over time, so we need to keep replicating
- And external validity can remain a problem
- All of this comes at enormous cost, and diverts attention from more helpful evaluation



Rethinking the Rush to Randomize: Another Problem—Spurious Results

- “Most research findings are false for most research designs and most fields”, including results of many RCTs (Ioannidis, 2005)
- Even a more limited analysis suggests that at least 14% of major published medical research is wrong (Jager & Leeks, 2013)
- Many problematic findings in published economics research (De Long & Lang, 1992) and many questionable findings in economics RCTs (Eble, Boone & Elbourne, 2014)



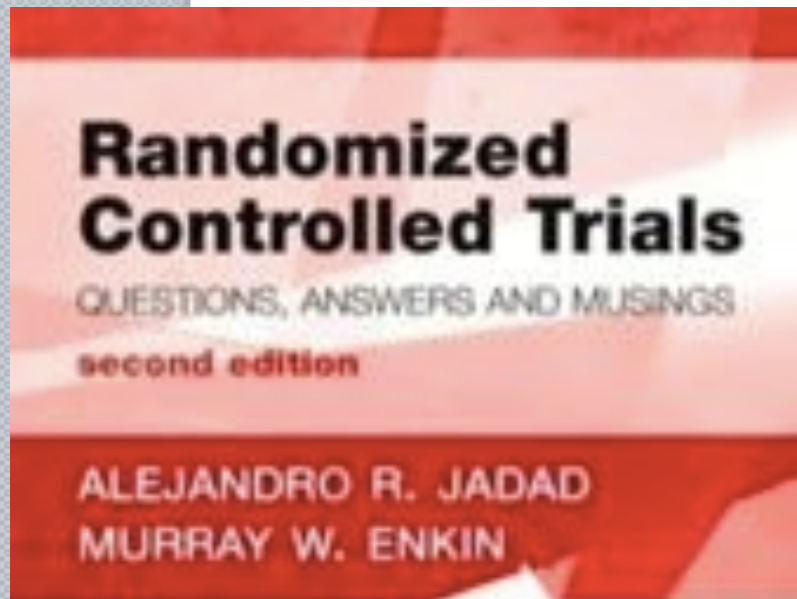
Rethinking the Rush to Randomize: Reasons for Spurious Results

- Underpowered studies – most social science and government RCTs are too small and lack large effect sizes
- Lack of replications – most social science and government RCTs lack sufficient replication
- Bias persists in most social science and government research designs, including RCTs, and can result in under or overestimation of effects sizes

Rethinking the Rush to Randomize: Bias Can Distort Claimed Effects Sizes

- Numerous commentators describe the bias that too often plagues research designs, including RCTs (see most recently Jadad and Enkin, 2014):

- ✓ Ascertainment bias
- ✓ Publication bias
- ✓ Selection bias
- ✓ Choice of question bias
- ✓ Wrong design bias
- ✓ Outcome choice bias
- ✓ Intervention choice bias
- ✓ Population choice bias
- ✓ Control group bias
- ✓ Attrition bias
- ✓ Selective results reporting bias
- ✓ Intentional fraud bias
- ✓ Time lag bias
- ✓ Regulation bias





Rethinking the Rush to Randomize

**What's
to Be Done?**



Rethinking the Rush to Randomize: Recommendations

- AEA should reengage with OMB and the White House on the uses and misuses of evaluation
- AEA also needs to engage with think tanks and the Congress
- Build out multiple methods expertise in federal evaluation capacity – not just RCT expertise
- Curtail use of RCTs for program effectiveness ratings – it's not meaningful
- Replace “what works” with “best practice” clearinghouses based on systematic reviews
- Refocus federal evaluation on program improvement instead of merely providing fodder for budget cuts