

# **Promoting transparency in Economics: Prospects for pre-specifying research hypotheses**

Professor Ted Miguel  
Department of Economics, U.C. Berkeley

*Stanford METRICS Forum*

May 16, 2016

- **The Berkeley Initiative for Transparency in the Social Sciences (BITSS)**
- <http://www.bitss.org/>



# The starting point

- There are numerous **problems** with empirical social science research, including: literatures distorted by false-positive results (Ioannidis 2005), data mining and tendentious reporting (Card and Krueger 1995), and large numbers of null findings that are “invisible” to the research community (Franco et al. 2014)
- There are also new tools for **diagnosing** and quantifying these problems (i.e., Simonsohn et al 2015 p-curve)
- But how to **solve** these problems?

# An overview of solutions

- Miguel et al (2014) lays out three inter-related approaches to begin addressing these problems:
  1. Disclosure (Simmons et al 2011)
  2. Open data and materials (→ replication)(But how to verify that the claims “disclosed” by authors are correct, and that the data shared is complete?)

## **3. Pre-registration of research hypotheses**

- Sharing research design and plans beforehand makes these other two approaches more useful, and has a range of other potential benefits.
- Replication may not be “enough”, i.e., context.

# (1) Why pre-specify research hypotheses?

- First of all, what is pre-registration?
- A researcher posts her/his research hypotheses, the data used to test them, and the planned research design (i.e., methodology) in a publicly available registry
- There is obviously a wide range of detail one could potentially include in an analysis plan (both [clinicaltrials.gov](https://clinicaltrials.gov) and the new AEA registry allow a researcher to include relatively sparse information)
- It is worth exploring the required fields in the AEA registry in a little more detail


# (1) Why pre-specify research hypotheses?

- The American Economics Association (AEA) registry, **[socialscienceregistry.org](https://socialscienceregistry.org)**, was founded in May 2013 with a focus on randomized control trials (RCTs).

Google News x Google Calendar x Google News x Golden State Files x clinicaltrials.gov x AEA RCT Registry x how do you print x

← → ↻ 🏠 <https://www.socialscienceregistry.org/site/about> 🔍 ☆ ☰

Apps Google News Google Gmail: Email from G... Google Calendar Other bookmarks



# AEA RCT Registry

The American Economic Association's registry for randomized controlled trials

[Create Account](#) [Sign in](#)

[About RCTs](#) [Registration Guidelines](#) [FAQ](#)

Advanced Search  [SEARCH](#)

## ABOUT THE REGISTRY

[REGISTER A TRIAL >](#)

---

### Welcome.

This is the American Economic Association's registry for randomized controlled trials.

Randomized Controlled Trials (RCTs) are widely used in various fields of economics and other social sciences. As they become more numerous, a central registry on which trials are on-going or complete (or abandoned) becomes important for various reasons: as a source of results for meta-analysis; as a one-stop resource to find out about available survey instruments and data.

Because existing registries are not well suited to the need for social sciences, in April 2012, the AEA executive committee decided to establish such a registry for economics and other social sciences.

**If you are running or have run a trial:** Registration is free and you do not need to be a member of the AEA to register. We encourage you to register any new study at its outset. However, given the backlog of existing trials, we invite you to also register past studies.

**If you are searching for results:** Please browse the data base. More results are forthcoming!

© Copyright 2012-2015, MIT.

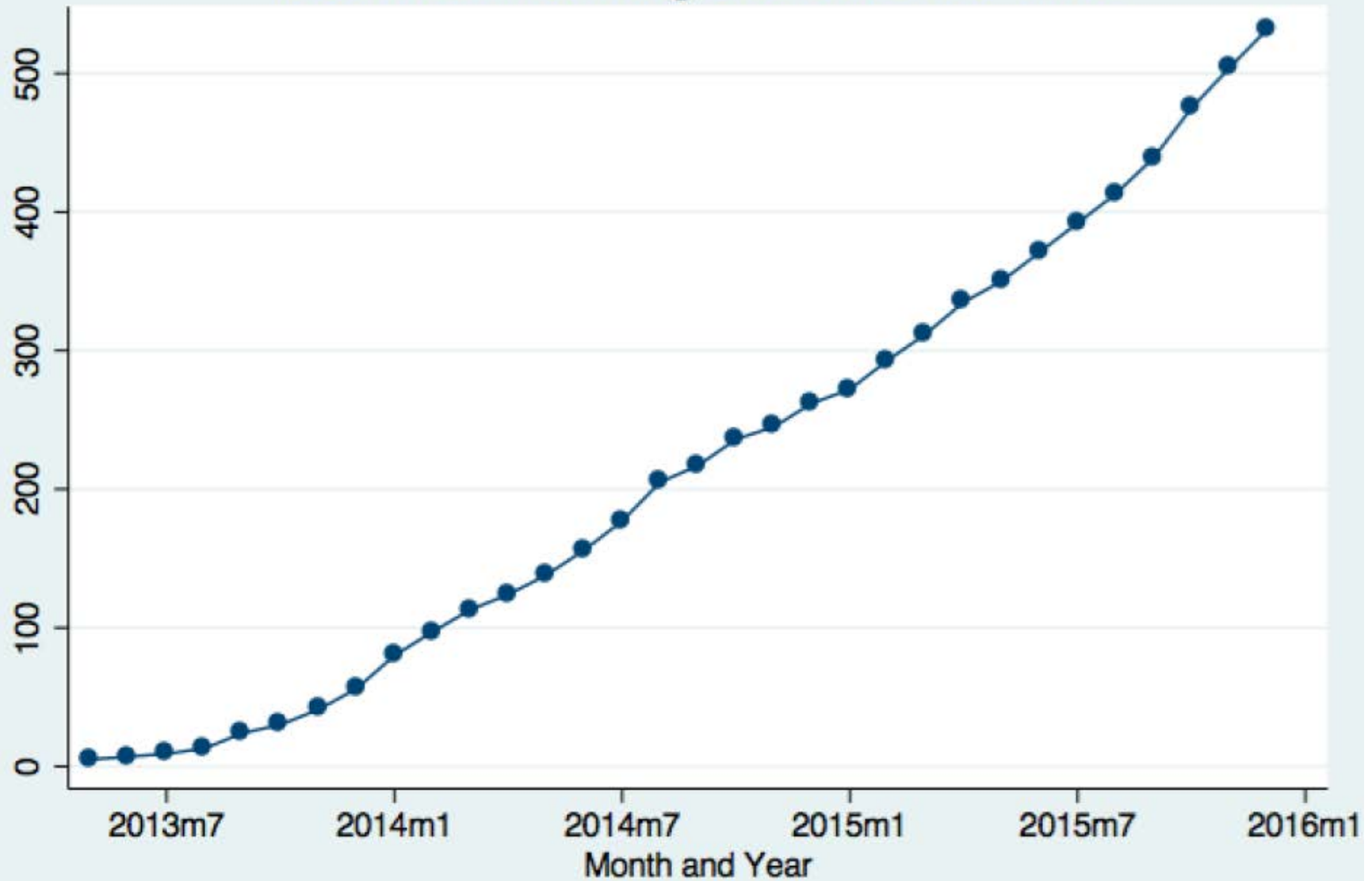
[About](#) [FAQ](#) [Contact](#)

# (1) Why pre-specify research hypotheses?

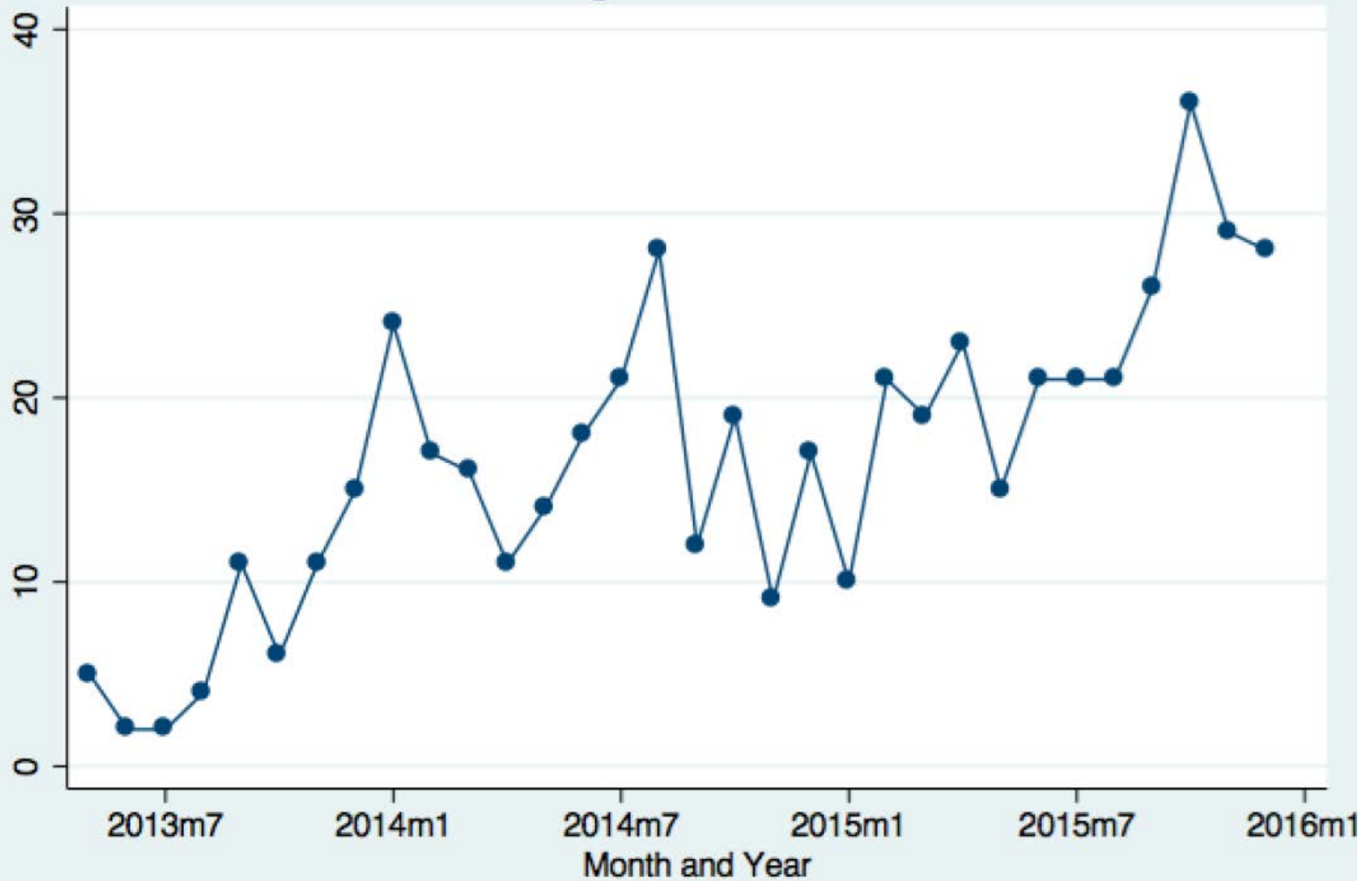
- The American Economics Association (AEA) registry, **[socialscienceregistry.org](http://socialscienceregistry.org)**, was founded in May 2013 with a focus on randomized control trials (RCTs).
- Since then over 660 studies have been registered, and the numbers are increasing rapidly.
- Some of these are earlier projects that are being registered (for completeness), but most are new studies.



## Total AEA Trial Registrations over Time



## New AEA Registrations Each Month



# (1) Why pre-specify research hypotheses?

- The required information on the AEA site includes:
- Trial Title; Country; Status (i.e., ongoing, completed); Keywords; **Abstract**; Trial Start Date; Intervention Start Date; Intervention End Date; Trial End Date; **Outcomes (End Points)**; **Experimental Design (Public)**; Was the treatment clustered?; Planned Number of Clusters; Planned Number of Observations; IRB approval info.

# (1) Why pre-specify research hypotheses?

The screenshot shows a web browser window with two tabs labeled 'AEA RCT Registry'. The address bar displays the URL <https://www.socialscienceregistry.org/trials/new>. The browser's toolbar includes navigation buttons (back, forward, refresh, home), a search bar, and a list of bookmarks (Apps, Google News, Google, Gmail, Google Calendar). The website header features the AEA RCT Registry logo and the text 'The American Economic Association's registry for randomized controlled trials'. A welcome message 'Welcome back, Edward!' is displayed, along with links for 'Sign out', 'Profile', and 'My Trials'. A navigation bar contains links for 'About RCTs', 'Registration Guidelines', and 'FAQ', as well as an 'Advanced Search' section with a search input field and a 'SEARCH' button. The main content area is titled 'NEW TRIAL INFORMATION' and contains a form with the following fields: 'Trial Title \*' (a text input field), 'Status \*' (a label), 'Study Status' (a dropdown menu), and a 'CREATE TRIAL' button. The footer of the website includes the copyright notice '© Copyright 2012-2015, MIT.' and links for 'About', 'FAQ', and 'Contact'. The Windows taskbar at the bottom shows various application icons, including the Start button, Chrome, Firefox, and several office applications, along with system status icons (network, volume, battery) and the system clock showing '9:27 AM 2/10/2015'.

**AEA RCT Registry**  
The American Economic Association's registry for randomized controlled trials

Welcome back, Edward!  
Sign out Profile My Trials

About RCTs Registration Guidelines FAQ Advanced Search  **SEARCH**

## NEW TRIAL INFORMATION

**Trial Title \***

**Status \***  
Study Status

**CREATE TRIAL**

© Copyright 2012-2015, MIT. About FAQ Contact

# (1) Why pre-specify research hypotheses?

- **Why might pre-registration be useful?**
  1. **Rounds out the body of evidence** by creating a “paper trail” of unpublished studies in an area → potentially helping to address publication bias (similar to Franco et al 2014) and improve meta-analysis.

# (1) Why pre-specify research hypotheses?

- **Why might pre-registration be useful?**
  1. **Rounds out the body of evidence** by creating a “paper trail” of unpublished studies in an area
  2. **Reduces the risk of data mining** and other tendentious presentation of results (“cherry-picking”) → by making clear what the authors’ original intentions and research hypotheses actually were.

# (1) Why pre-specify research hypotheses?

- **Why might pre-registration be useful?**
  - 1. Rounds out the body of evidence** by creating a “paper trail” of unpublished studies in an area
  - 2. Reduces the risk of data mining** and other tendentious presentation of results (“cherry-picking”)
  - 3. Generates correctly sized statistical tests**, bolstering the credibility of statistical significance levels → by making clear what additional tests were run beyond those originally planned, and thus making multiple testing adjustments more credible.

# (1) Why pre-specify research hypotheses?

- **Why might pre-registration be useful?**
  1. **Rounds out the body of evidence** by creating a “paper trail” of unpublished studies in an area
  2. **Reduces the risk of data mining** and other tendentious presentation of results (“cherry-picking”)
  3. **Generates correctly sized statistical tests**, bolstering the credibility of statistical significance levels
  4. **Makes open data and disclosure more effective** → by allowing other scholars to cross-check published information against original research plans.



# (1) Why pre-specify research hypotheses?

- **Why might pre-registration be useful?**
  1. **Rounds out the body of evidence** by creating a “paper trail” of unpublished studies in an area
  2. **Reduces the risk of data mining** and other tendentious presentation of results (“cherry-picking”)
  3. **Generates correctly sized statistical tests**, bolstering the credibility of statistical significance levels
  4. **Makes open data and disclosure more effective**
  5. As a side benefit, **forces researchers to more carefully think through their hypotheses beforehand, improving research quality** → reducing “waste” of funding on poorly conceived projects

# (1) Why pre-specify research hypotheses?

- **Why might pre-registration be useful?**

1. Rounds out the body of evidence
2. Reduces the risk of data mining
3. Generates correctly sized statistical tests
4. Makes open data and disclosure more effective
5. Forces researchers to more carefully think through their hypotheses beforehand, improving research quality
6. **Others?**

# (1) Why pre-specify research hypotheses?

- A leading concern: will pre-registration of plans **stifle creativity** and limit discoveries made through exploratory research?
- Many, if not most, important scientific findings undoubtedly originated as unexpected discoveries...

# (1) Why pre-specify research hypotheses?

- A leading concern: will pre-registration of plans **stifle creativity** and limit discoveries made through exploratory research?
- Many, if not most, important scientific findings undoubtedly originated as unexpected discoveries...
- But findings from such work are inherently more tentative because of the greater flexibility of tests and the greater opportunity for the outcome to obtain by chance.  
→ Pre-specification is not intended to disparage exploratory analysis, but rather to **free it from the tradition of being portrayed as formal hypothesis testing.**

## (2) **How widely** can pre-specification be applied?

- A major open intellectual question (and the focus of today's lecture)

## (2) **How widely** can pre-specification be applied?

- A major open intellectual question (and the focus of today's lecture)

**1. Laboratory experiments:** pre-analysis plans could be particularly important given the relatively low cost to researchers of running multiple experiments and never publishing the results from those that “didn't work”

- Seems like low-hanging fruit
- The flip-side is that lab experiments should also be relatively low cost to replicate (Niederle 2015)

## (2) **How widely** can pre-specification be applied?

### 2. **Prospective observational (non-experimental)**

**studies\***: One promising area is the registration of studies of anticipated policy changes.

- First pre-analysis plan in Economics (to my knowledge) was Neumark's (1999, 2001) plan to study the effect of future minimum wage increases on unemployment – discussed next.

## (2) **How widely** can pre-specification be applied?

- Pre-registration can also be used when new “rounds” of **data are released** (e.g., a new PSID wave, Census round), or where access to existing data is restricted and thus where data mining is impossible ex ante.
- (What share of quantitative empirical work?)
- The **Open Science Framework** (OSF) provides a flexible platform for time-stamping and archiving materials to be made publicly available.



## (2) **How widely** can pre-specification be applied?

- **Beyond applied micro empirical studies:** To reduce concerns about “specification search”, researchers could also pre-register:
  - the parameters to be used in macroeconomic calibrations (“quantitative exercises”),
  - the models used in structural estimation (i.e., in industrial organization), Bai et al (2015),
  - prior distributions used in Bayesian analysis (perhaps gathered through eliciting expert opinion).
- Plans archived but only published with a time lag (on the AEA registry or OSF site) to make sure researchers with creative ideas are not “scooped” by others.

### (3) Neumark (1999, 2001)

- This paper is a (largely forgotten) **milestone in social science** research methodology
- To my knowledge, **first pre-analysis plan** in Economics
- Study of the highly contentious (and politicized) issue of labor market impacts of minimum wage increases
- Card and Krueger's (1995) point about publication bias in this area is a starting point

### (3) Neumark (1999, 2001)

TABLE 1

ELASTICITIES OF EMPLOYMENT WITH RESPECT TO THE MINIMUM WAGE  
FROM RECENT STUDIES<sup>a</sup>

Study	Data source	Workers studied	Elasticities
Card and Krueger (1994)	Survey of fast-food restaurants in NJ and PA	Fast-food workers	0.61 (*), 0.69 (**) <sup>b</sup>
Card (1992a)	1989–1990 CPS Outgoing Rotation Group files, April–December	Teenagers	~0 <sup>c</sup>
Card (1992b)	1987–1989 CPS Outgoing Rotation Group files, for CA and five states or cities as comparison groups	Teenagers	0.47 (**) <sup>d</sup>
Katz and Krueger (1992)	Survey of fast-food restaurants in TX	Fast-food workers	1.73 (*), 1.85 (*) <sup>e</sup>
Neumark and Wascher (1992, 1994)	May CPS files, 1973–1989	Teenagers, young adults aged 16–24	~0 (teenagers, without enrollment control), –0.1 to –0.2 (*) (teenagers, with enrollment control), –0.15 to –0.2 (**) (young adults, with or without enrollment control) <sup>f</sup>
Card, Katz, and Krueger (1994)	Data from Neumark and Wascher (1992)	Teenagers	0.37 (**) <sup>g</sup>
Kim and Taylor (1995)	County Business Patterns, March 1988–March 1989, CA and rest of U.S.	Retail-trade workers	<0 (**) <sup>h</sup>
Deere, Murphy, and Welch (1995)	CPS Outgoing Rotation Group files, 1985–1993	Teenagers (15–19)	–0.27 (males), –0.42 (**) (females), –0.37 (**) (blacks) <sup>i</sup>
Currie and Fallick (1996)	NLSY	Youths	–0.09 (**) <sup>j</sup>
Neumark and Wascher (1998a)	Resurvey of fast-food restaurants in NJ and PA	Fast-food workers	–0.1 to –0.3 (**) <sup>k</sup>

Considerable  
variation in  
estimated effects

### (3) Neumark (1999, 2001)

- Solution (in the absence of a study registry): submitted a detailed analysis plan to a journal (*Industrial Relations*) before the release of the relevant U.S. government employment statistics (in May 1997)
- Received referee reports and revised the analysis plan **before receiving the data**

### (3) Neumark (1999, 2001)

Given this latter possibility, the goal of this symposium, as I understand it, was to attempt to obtain estimates of employment effects of minimum wages that are free from any bias that might be introduced via the process of specification search. This was to be accomplished by proposing that various authors involved in the minimum wage debate (1) prespecify a research design that would use data not yet available, (2) subject that research design to peer review before any data analysis was performed, and (3) carry out the prespecified analysis when the data became available. The evidence resulting from such prespecified research designs should provide estimates of the employment effects of minimum wages that are free of author effects by eliminating specification search.

(p. 123)

### (3) Neumark (1999, 2001)

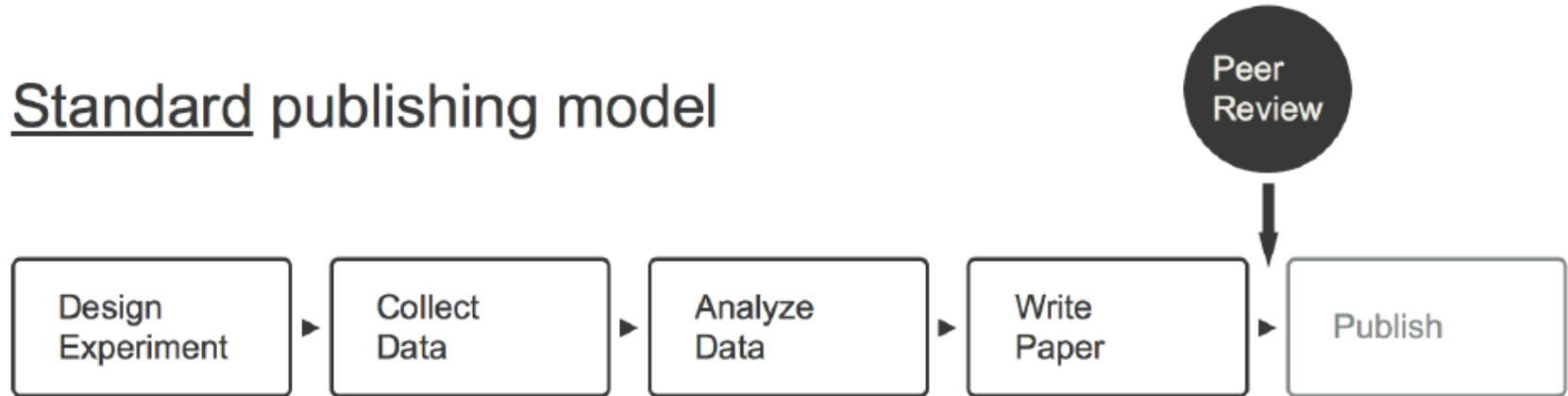
- Solution (in the absence of a study registry): submitted a detailed analysis plan to a journal (*Industrial Relations*) before the release of the relevant U.S. government employment statistics (in May 1997)
- Received referee reports and revised the analysis plan **before receiving the data**
- Similar to the idea of “**adversarial collaboration**”, often associated with Daniel Kahneman

### (3) Neumark (1999, 2001)

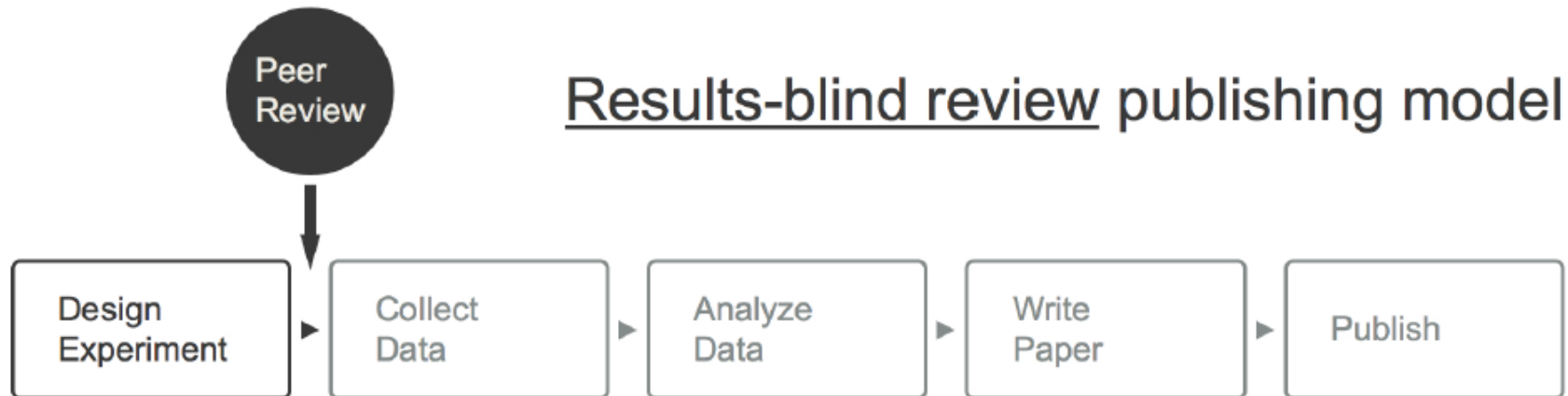
- Also very close to a “**registered report**”, a format currently used in some psychology and scientific journals: paper basically “accepted” based on introduction, literature review and analysis plan, before generating any results

# The Results-Blind Review Publishing Model for Registered Reports (RRs)

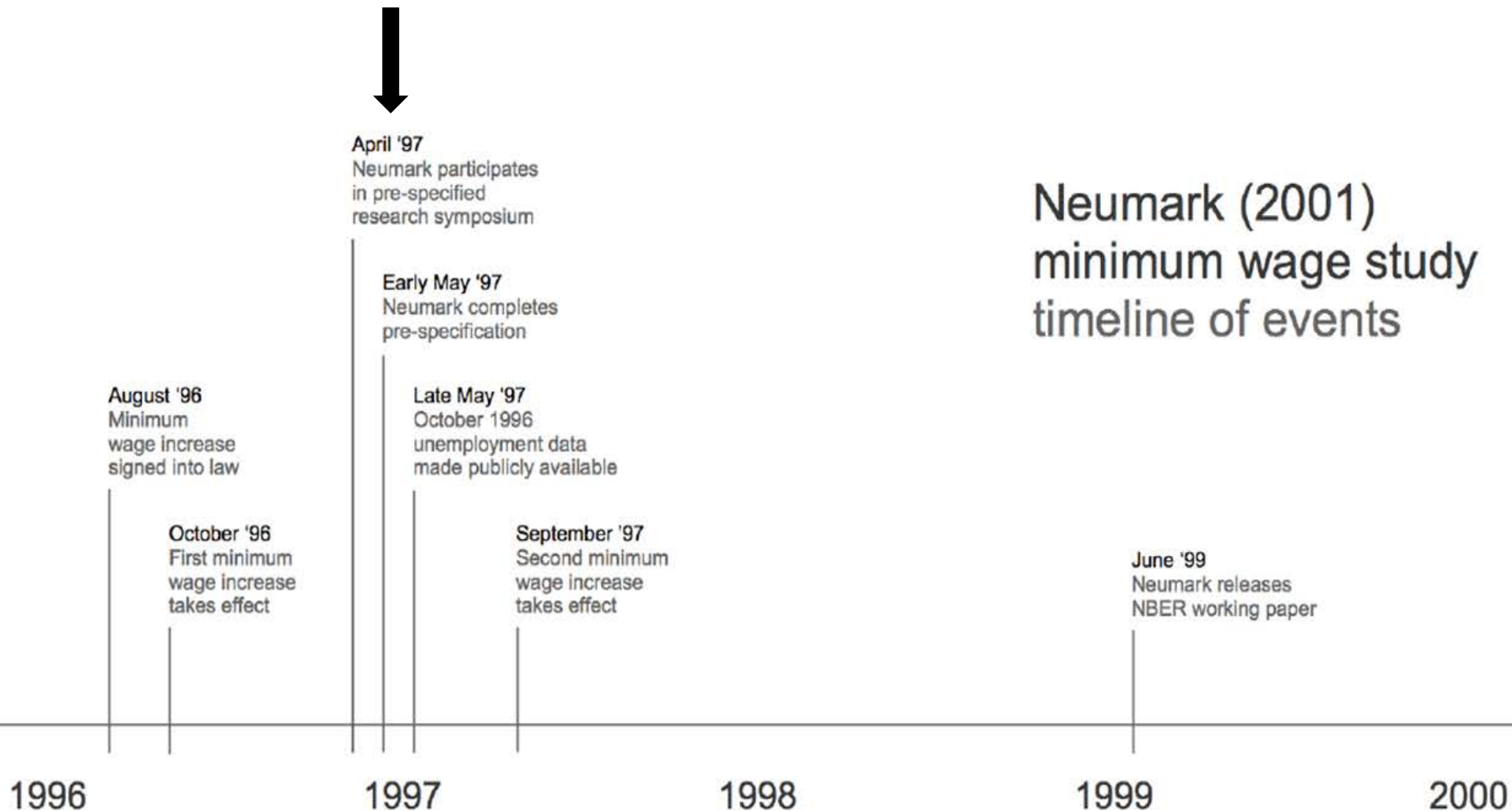
## Standard publishing model



## Results-blind review publishing model







### (3) Neumark (1999, 2001)

- An interesting and important point about the possible costs of pre-specification (p. 133): the advantages of tying researchers' hands (at least partially) need to be weighed against **the loss of “specification analysis”**, i.e., using an appropriate statistical model
- Choosing a model can be difficult ex ante, since it may depend on the variation in the (yet unobserved) data

### (3) Neumark (1999, 2001)

- He argues that this concern is less important in **well-established areas of inquiry** such as the study of minimum wage impacts where the data is widely used and well-understood

### (3) Neumark (1999, 2001)

sively. As is often the case in applied econometrics, removing one source of bias may exacerbate other sources of bias and hence does not necessarily leave us with less biased estimates.<sup>16</sup> In particular, because economic theory rarely provides us with the functional forms, lag structures, etc. that should be used in empirical work, careful specification analysis plays an important role. While the prespecified research design approach should eliminate bias from specification search, it also may be prone to errors stemming from the absence of specification analysis. On the other hand, the prespecified research design approach may be most useful in areas of inquiry—such as employment effects of minimum wages—in which researchers have in the past thought carefully, based on both theory and existing evidence, about the specifications to be estimated. In such areas, prespecified designs are less likely to include severe specification errors that should be corrected based on examination of the data. With these provisos in mind, the next section reports the estimates.

(p. 133)

### (3) Neumark (1999, 2001)

- He argues that this concern is less important in **well-established areas of inquiry** such as the study of minimum wage impacts where the data is widely used and well-understood
- The flip-side is that there is inherently less leeway to data mine in such a research area, since there is already some consensus regarding the statistical approach (i.e., referees will want to see a “standard specification”)
- More speculatively, in areas where there is no standard approach (i.e., in Casey et al’s 2012 study of community driven development programs in Sierra Leone) pre-specification may be equally or more valuable

### (3) Neumark (1999, 2001)

TABLE 3  
ESTIMATED EFFECTS OF MINIMUM WAGES ON EMPLOYMENT OF  
16–19-YEAR-OLDS<sup>a</sup>

	(1)	(2)	(3)
Fixed state and year effects, basic files:			
Relative minimum wage	0.098 (0.241)	—	0.099 (0.241)
Relative minimum wage, lagged one year	−0.025 (0.246)	—	−0.015 (0.247)
Elasticity	0.064	—	0.073

→ What did Neumark find? No statistically significant effects and small magnitudes, somewhere in middle of existing estimates. But analysis appears under-powered (large SE's)

### (3) Neumark (1999, 2001)

- An important decision: Neumark denotes any additional non-prespecified analysis in the final version of the text:

Below the basic specification, results from some alternatives are reported. First, estimates are reported when only the lagged minimum wage variable is used.<sup>18</sup> The point estimates are still very small and

---

<sup>18</sup> Although this particular restriction was not mentioned in the second section of this article, in some of the subsequent tables the results point to negative, statistically significant lagged employment effects and positive but insignificant contemporaneous effects, raising the possibility that the lagged effects would not be significant if considered in isolation. Estimates of the restricted model address this question.

- Is this an appropriate specification to focus on? A judgment call

### (3) Neumark (1999, 2001)

- A main limitation of this approach, as articulated by Neumark, is a lack of statistical precision:
- *“Because such an approach has to rely on relatively little data (unless one commits to the design a decade or more in advance), it may be difficult to draw strong conclusions of any kind, even when the point estimates are consistent with past research in which similar magnitudes were statistically significant.”* (pp. 136-137)
- (But why not just keep using this specification and enrich the analysis with more data as it comes in over time? E.g., many minimum wage policy changes will occur.)



### (3) Neumark (1999, 2001)

- Beyond a simple registered pre-analysis plan, the approach is closer to a **registered report**, which is an even more decisive break with current journal publication practices
- My bottom line: there are many potential social science applications of this approach related to the public release of new data; anticipated policy changes; future election results, etc.
- Any additional comments?

# Research transparency

- **The Berkeley Initiative for Transparency in the Social Sciences (BITSS)** is working to promote solutions:
  - Training courses, workshops
  - Online pedagogical materials, “how to” guides
  - Grant competitions on meta-research, replication
  - Prizes to recognize younger leaders in the field
  - Funding efforts in our “catalyst network”
- Check out <http://www.bitss.org/>



## (4) Dal-Re et al (2014): Registering OS's

## (4) Dal-Re et al (2014): Registering OS's

- While there remain a number of open questions and challenges going forward, at least conceptually, the pre-registration of field experiments (Casey et al 2012), lab experiments, and **prospective** non-experimental studies (Neumark 2001) seems straightforward to incorporate into current research practices in the social sciences
- What about other (non-experimental) **observational studies (OS's)**?
- Is it desirable for non-prospective non-experimental studies to be registered? How to do it?

## (4) Dal-Re et al (2014): Registering OS's

- Dal-Re et al. first show that OS's constitute the vast majority of human subjects based medical research studies published in 2011: around **90% of studies**, while fewer than 6% were RCT's
- The same is true in social science fields, e.g., Oster (2014) finds that **>85% of published empirical papers** in leading economics journals are non-experimental
- Big picture question: how can improved transparency practices be brought into the bulk of quantitative empirical research?

## (4) Dal-Re et al (2014): Registering OS's

- Currently, **no consensus** on the registration for OS's in medical research or epidemiology (dueling editorial statements in leading epidemiology journals in 2009-10)
- E.g., Editors, “The registration of observational studies—When metaphors go bad.” *Epidemiology* 21 (2010).

## (4) Dal-Re et al (2014): Registering OS's

- Currently, **no consensus** on the registration for OS's in medical research or epidemiology (dueling editorial statements in leading epidemiology journals in 2009-10)
- E.g., Editors, “The registration of observational studies—When metaphors go bad.” *Epidemiology* 21 (2010).
- Dal-Re et al make a strong call in favor. Why?
  1. If already get IRB approval, minimal additional burden
  2. Make the totality of the evidence “more visible” to other scholars, i.e., even unpublished studies
  3. Speculatively, might increase publication of null findings
  4. Other benefits?

## (4) Dal-Re et al (2014): Registering OS's

- Dal-Re et al argue against some concerns.
  1. A (the?) leading concern is that there is **no way to verify whether registration precedes analysis**, leading to a false sense of confidence in OS results (or even greater skepticism about their findings)



## (4) Dal-Re et al (2014): Registering OS's

### **THE REGISTRATION CHALLENGE**

In contrast to CTs, which require prospective data collection and follow-up of participants, analyses of some OSs can be performed readily in minimal time whenever required data have been collected, perhaps as part of a prior survey or a byproduct of health care activities (for example, administrative or billing databases, disease registries). In these cases, registering a protocol or a full analysis plan may not qualify as prospective. Theoretically, an investigator can mine the available data, notice some provocative results, and build a protocol and analysis plan around the selected results while spuriously claiming that the plan was prospectively conceived. Therefore, for ex-

(p. 2)

## (4) Dal-Re et al (2014): Registering OS's

- Dal-Re et al argue against some concerns.
  1. A (the?) leading concern is that there is **no way to verify whether registration precedes analysis**, leading to a false sense of confidence in OS results (or even greater skepticism about their findings)
  2. Other scholars could search a public registry for promising ideas using easily available data, and “scoop” the original authors. (Would most credit to those with the registered plan or those who publish first?)

## (4) Dal-Re et al (2014): Registering OS's

- Dal-Re et al argue against some concerns.
  1. A (the?) leading concern is that there is **no way to verify whether registration precedes analysis**, leading to a false sense of confidence in OS results (or even greater skepticism about their findings)
  2. Other scholars could search a public registry for promising ideas using easily available data, and “scoop” the original authors. (Would most credit to those with the registered plan or those who publish first?)
  3. Practically, developing pre-registration standards may be more difficult for OS's, given the wide range of methods and data they employ
  4. Other concerns?

## (4) Dal-Re et al (2014): Registering OS's

- Also make a strong case in favor of greater **data sharing**, including of anonymized individual-level data
- This practice is common in some social sciences (especially economics, political science) but much less so in medical research, perhaps due to privacy concerns
- In addition, support sharing full data “log” files of analysis or even “live streaming” of statistical analysis in real-time
- (But once again, the “honor system” – how to verify that this is truly prospective for publicly available data?)