Registrations and Pre-Analysis Plans Making research more transparent and reproducible

Garret Christensen¹

¹Berkeley Initiative for Transparency in the Social Sciences
University of California Berkeley
and
Center for Open Science

BITSS Annual Meeting, 2014



Outline

- Motivation
 - Publication Bias
 - P-Hacking
- Solutions
 - Registration
 - Pre-Analysis Plan
- 3 Conclusion

- There is a higher fraction of rejected hypothesis tests in the social sciences than in physical sciences (Fanelli 2010).
- Published null results are disappearing over time, in all disciplines (Fanelli 2011).
- This is very unlikely to represent the true state of the universe.
- Data on the complete set of experiments run shows strong results are 40pp more likely to be published, and 60pp more likely to be written up. The file drawer problem is massive. (Franco, Malhotra, Simonovits 2014—see tomorrow)



- There is a higher fraction of rejected hypothesis tests in the social sciences than in physical sciences (Fanelli 2010).
- Published null results are disappearing over time, in all disciplines (Fanelli 2011).
- This is very unlikely to represent the true state of the universe.
- Data on the complete set of experiments run shows strong results are 40pp more likely to be published, and 60pp more likely to be written up. The file drawer problem is massive. (Franco, Malhotra, Simonovits 2014—see tomorrow)



- There is a higher fraction of rejected hypothesis tests in the social sciences than in physical sciences (Fanelli 2010).
- Published null results are disappearing over time, in all disciplines (Fanelli 2011).
- This is very unlikely to represent the true state of the universe.
- Data on the complete set of experiments run shows strong results are 40pp more likely to be published, and 60pp more likely to be written up. The file drawer problem is massive. (Franco, Malhotra, Simonovits 2014—see tomorrow)



- There is a higher fraction of rejected hypothesis tests in the social sciences than in physical sciences (Fanelli 2010).
- Published null results are disappearing over time, in all disciplines (Fanelli 2011).
- This is very unlikely to represent the true state of the universe.
- Data on the complete set of experiments run shows strong results are 40pp more likely to be published, and 60pp more likely to be written up. The file drawer problem is massive. (Franco, Malhotra, Simonovits 2014—see tomorrow)

If we only write up/publish significant results, and we have no record of all the insignificant results, we have no way to tell if our 'significant' results are real, or if they're the 5% we should expect due to randomness.

- Not something only evil people do. It can be subconcious.
- Also called fishing, researcher degrees of freedom, data mining, data massaging, or specification searching.
- Definition: flexibility in data analysis allows portrayal of anything as below an arbitrary p-value threshhold; significance loses its meaning.

- Not something only evil people do. It can be subconcious.
- Also called fishing, researcher degrees of freedom, data mining, data massaging, or specification searching.
- Definition: flexibility in data analysis allows portrayal of anything as below an arbitrary p-value threshhold; significance loses its meaning.

- Not something only evil people do. It can be subconcious.
- Also called fishing, researcher degrees of freedom, data mining, data massaging, or specification searching.
- Definition: flexibility in data analysis allows portrayal of anything as below an arbitrary p-value threshhold; significance loses its meaning.

- Not something only evil people do. It can be subconcious.
- Also called fishing, researcher degrees of freedom, data mining, data massaging, or specification searching.
- Definition: flexibility in data analysis allows portrayal of anything as below an arbitrary p-value threshhold; significance loses its meaning.

study's dependent measures

2. Deciding whether to collect more data after looking to see whether the results were significant

I. In a paper, failing to report all of a

3. In a paper, failing to report all of a study's conditions
4. Stopping collecting data earlier than planned because one found the result that one had been

5. In a paper, "rounding off" a

looking for

Does this actually happen?

(JLP 2011)

- p value (e.g., reporting that a p value of .054 is less than .05)
 6. In a paper, selectively reporting studies that "worked"
 7. Deciding whether to exclude data
- after looking at the impact of doing so on the results
 8. In a paper, reporting an unexpected finding as having been predicted from the start
 9. In a paper, claiming that results
- 9. In a paper, claiming that results are unaffected by demographic variables (e.g., gender) when one is actually unsure (or knows that they do)

 10. Falsifying data

 In a paper, failing to report all of a study's dependent measures 	63.4
Deciding whether to collect more data after looking to see whether the results were significant	55.9
In a paper, failing to report all of a study's conditions	27.7
 Stopping collecting data earlier than planned because one found the result that one had been looking for 	15.6
 In a paper, "rounding off" a p value (e.g., reporting that a p value of .054 is less than .05) 	22.0
In a paper, selectively reporting studies that "worked"	45.8
 Deciding whether to exclude data after looking at the impact of do- ing so on the results 	38.2
 In a paper, reporting an unex- pected finding as having been predicted from the start 	27.0
 In a paper, claiming that results are unaffected by demographic variables (e.g., gender) when one is actually unsure (or knows that they do) 	3.0
10. Falsifying data	0.6

	Admission rate	Defensibility rate
I. In a paper, failing to report all of a study's dependent measures	63.4	1.84 (0.39)
 Deciding whether to collect more data after looking to see whether the results were significant 	55.9	1.79 (0.44)
In a paper, failing to report all of a study's conditions	27.7	1.77 (0.49)
 Stopping collecting data earlier than planned because one found the result that one had been looking for 	15.6	1.76 (0.48)
 In a paper, "rounding off" a p value (e.g., reporting that a p value of .054 is less than .05) 	22.0	1.68 (0.57)
In a paper, selectively reporting studies that "worked"	45.8	1.66 (0.53)
 Deciding whether to exclude data after looking at the impact of do- ing so on the results 	38.2	1.61 (0.59)
 In a paper, reporting an unex- pected finding as having been predicted from the start 	27.0	1.50 (0.60)
 In a paper, claiming that results are unaffected by demographic variables (e.g., gender) when one is actually unsure (or knows that they do) 	3.0	1.32 (0.60)
10. Falsifying data	0.6	0.16 (0.38)

- Do people actually do this? (Previous—John, Loewenstein, Prelec 2011)
- Listening to the Beatles' "When I'm Sixty-Four" makes you younger. (Simmons, Nelson, Simonsohn 2011)
- Inordinately many .049 p-values, and indordinately few .051's. (Brodeur et al 2013)
- Political idiologues literally see in black and white (Nosek, Spies, Motyl 2012)

- Do people actually do this? (Previous—John, Loewenstein, Prelec 2011)
- Listening to the Beatles' "When I'm Sixty-Four" makes you younger. (Simmons, Nelson, Simonsohn 2011)
- Inordinately many .049 p-values, and indordinately few .051's. (Brodeur et al 2013)
- Political idiologues literally see in black and white (Nosek, Spies, Motyl 2012)

- Do people actually do this? (Previous—John, Loewenstein, Prelec 2011)
- Listening to the Beatles' "When I'm Sixty-Four" makes you younger. (Simmons, Nelson, Simonsohn 2011)
- Inordinately many .049 p-values, and indordinately few .051's. (Brodeur et al 2013)
- Political idiologues literally see in black and white (Nosek, Spies, Motyl 2012)

- Do people actually do this? (Previous—John, Loewenstein, Prelec 2011)
- Listening to the Beatles' "When I'm Sixty-Four" makes you younger. (Simmons, Nelson, Simonsohn 2011)
- Inordinately many .049 p-values, and indordinately few .051's. (Brodeur et al 2013)
- Political idiologues literally see in black and white (Nosek, Spies, Motyl 2012)

Solutions

- Publication Bias—Registration
- P-Hacking—Pre-Analysis Plans

Solutions

- Publication Bias—Registration
- P-Hacking—Pre-Analysis Plans

- Publicly stating all research you do, what hypotheses you test, prospectively.
- Near universal adoption in medical RCTs. http://clinicaltrials.gov
- Newer to social sciences, but:
 - AEA registry http://socialscienceregistry.org
 - EGAP registry
 - http://egap.org/design-registration
 - 3ie registry http://ridie.3ieimpact.org
 - Open Science Framework http://osf.io
 - Open format
 - Will soon sync with above



- Publicly stating all research you do, what hypotheses you test, prospectively.
- Near universal adoption in medical RCTs.

```
http://clinicaltrials.gov
```

- Newer to social sciences, but:
 - AEA registry http://socialscienceregistry.org
 - EGAP registry
 - http://egap.org/design-registration
 - 3ie registry http://ridie.3ieimpact.org
 - Open Science Framework http://osf.io

- Publicly stating all research you do, what hypotheses you test, prospectively.
- Near universal adoption in medical RCTs. http://clinicaltrials.gov
- Newer to social sciences, but:
 - AEA registry http://socialscienceregistry.org
 - EGAP registry http://egap.org/design-registration
 - 3ie registry http://ridie.3ieimpact.org
 - Open Science Framework http://osf.io
 - Open format
 - Will soon sync with above



- Publicly stating all research you do, what hypotheses you test, prospectively.
- Near universal adoption in medical RCTs.

```
http://clinicaltrials.gov
```

- Newer to social sciences, but:
 - AEA registry http://socialscienceregistry.org
 - EGAP registry http://egap.org/design-registration
 - 3ie registry http://ridie.3ieimpact.org
 - Open Science Framework http://osf.io
 - Open format
 - Will soon sync with above



- Publicly stating all research you do, what hypotheses you test, prospectively.
- Near universal adoption in medical RCTs.

```
http://clinicaltrials.gov
```

- Newer to social sciences, but:
 - AEA registry http://socialscienceregistry.org
 - EGAP registry

```
http://egap.org/design-registration
```

- 3ie registry http://ridie.3ieimpact.org
- Open Science Framework http://osf.io
 - Open format
 - Will soon sync with above



- Publicly stating all research you do, what hypotheses you test, prospectively.
- Near universal adoption in medical RCTs.

```
http://clinicaltrials.gov
```

- Newer to social sciences, but:
 - AEA registry http://socialscienceregistry.org
 - EGAP registry

```
http://egap.org/design-registration
```

- 3ie registry http://ridie.3ieimpact.org
- Open Science Framework http://osf.io
 - Open format
 - Will soon sync with above



- Publicly stating all research you do, what hypotheses you test, prospectively.
- Near universal adoption in medical RCTs.

```
http://clinicaltrials.gov
```

- Newer to social sciences, but:
 - AEA registry http://socialscienceregistry.org
 - EGAP registry

```
http://egap.org/design-registration
```

- 3ie registry http://ridie.3ieimpact.org
- Open Science Framework http://osf.io
 - Open format
 - Will soon sync with above



- Publicly stating all research you do, what hypotheses you test, prospectively.
- Near universal adoption in medical RCTs.

```
http://clinicaltrials.gov
```

- Newer to social sciences, but:
 - AEA registry http://socialscienceregistry.org
 - EGAP registry

```
http://egap.org/design-registration
```

- 3ie registry http://ridie.3ieimpact.org
- Open Science Framework http://osf.io
 - Open format
 - Will soon sync with above



- Publicly stating all research you do, what hypotheses you test, prospectively.
- Near universal adoption in medical RCTs.

```
http://clinicaltrials.gov
```

- Newer to social sciences, but:
 - AEA registry http://socialscienceregistry.org
 - EGAP registry

```
http://egap.org/design-registration
```

- 3ie registry http://ridie.3ieimpact.org
- Open Science Framework http://osf.io
 - Open format
 - Will soon sync with above



Pre-Analysis Plan

- Often part of a registration

Pre-Analysis Plan

Often part of a registration

From 3ie: "A pre-analysis plan is a detailed description of

Pre-Analysis Plan

- Often part of a registration
- From 3ie: "A pre-analysis plan is a detailed description of the analysis to be conducted that is written in advance of seeing the data on impacts of the program being evaluated. It may specify hypotheses to be tested, variable construction, equations to be estimated, controls to be used, and other aspects of the analysis. A key function of the pre-analysis plan is to increase transparency in the research. By setting out the details in advance of what will be done and before knowing the results, the plan guards against data mining and specification searching. Researchers are encouraged to develop and upload such a plan with their study registration, but it is not required for registration."

- the main outcome measures,
- which outcome measures are primary and which are secondary,
- the precise composition of any families that will be used for mean effects analysis.
- the subgroups that will be analyzed,
- the direction of expected impact if we want to use a one-sided test, and
- the primary specification to be used for the analysis

- the main outcome measures,
- which outcome measures are primary and which are secondary,
- the precise composition of any families that will be used for mean effects analysis,
- the subgroups that will be analyzed,
- the direction of expected impact if we want to use a one-sided test, and
- 6 the primary specification to be used for the analysis.



- 1 the main outcome measures,
- which outcome measures are primary and which are secondary,
- the precise composition of any families that will be used for mean effects analysis,
- the subgroups that will be analyzed,
- the direction of expected impact if we want to use a one-sided test, and
- 6 the primary specification to be used for the analysis.



- the main outcome measures,
- which outcome measures are primary and which are secondary,
- the precise composition of any families that will be used for mean effects analysis,
- the subgroups that will be analyzed,
- the direction of expected impact if we want to use a one-sided test, and
- the primary specification to be used for the analysis



Glennerster, Takavarasha Suggestions

- the main outcome measures,
- which outcome measures are primary and which are secondary,
- the precise composition of any families that will be used for mean effects analysis,
- the subgroups that will be analyzed,
- the direction of expected impact if we want to use a one-sided test, and
- the primary specification to be used for the analysis.



Glennerster, Takavarasha Suggestions

- 1 the main outcome measures,
- which outcome measures are primary and which are secondary,
- the precise composition of any families that will be used for mean effects analysis,
- the subgroups that will be analyzed,
- the direction of expected impact if we want to use a one-sided test, and
- the primary specification to be used for the analysis.



Glennerster, Takavarasha Suggestions

- 1 the main outcome measures,
- which outcome measures are primary and which are secondary,
- the precise composition of any families that will be used for mean effects analysis,
- the subgroups that will be analyzed,
- the direction of expected impact if we want to use a one-sided test, and
- the primary specification to be used for the analysis.

- Description of the sample to be used in the study
- Mey data sources
- Hypotheses to be tested throughout the causal chain
- Specify how variables will be constructed
- Specify the treatment effect equation to be estimated
- What is the plan for how to deal with multiple outcomes and multiple hypothesis testing?
- Procedures to be used for addressing survey attrition
- How will the study deal with outcomes with limited variation?
- If you are going to be testing a model, include the mode
 - Remember to archive it



- Description of the sample to be used in the study
- Key data sources
- 4 Hypotheses to be tested throughout the causal chair
- Specify how variables will be constructed
- Specify the treatment effect equation to be estimated
- What is the plan for how to deal with multiple outcomes and multiple hypothesis testing?
- Procedures to be used for addressing survey attrition
- 4 How will the study deal with outcomes with limited variation?
- If you are going to be testing a model, include the model
- Remember to archive it



- Description of the sample to be used in the study
- Key data sources
- 4 Hypotheses to be tested throughout the causal chain
- Specify how variables will be constructed
- Specify the treatment effect equation to be estimated
- What is the plan for how to deal with multiple outcomes and multiple hypothesis testing?
- Procedures to be used for addressing survey attrition
- How will the study deal with outcomes with limited variation?
- If you are going to be testing a model, include the model
- Remember to archive it



- Description of the sample to be used in the study
- Key data sources
- Hypotheses to be tested throughout the causal chain
- Specify how variables will be constructed
- Specify the treatment effect equation to be estimated
- What is the plan for how to deal with multiple outcomes and multiple hypothesis testing?
- Procedures to be used for addressing survey attrition
- 4 How will the study deal with outcomes with limited variation?
- If you are going to be testing a model, include the model
- Remember to archive it



- Description of the sample to be used in the study
- Key data sources
- Hypotheses to be tested throughout the causal chain
- Specify how variables will be constructed
- Specify the treatment effect equation to be estimated
- What is the plan for how to deal with multiple outcomes and multiple hypothesis testing?
- Procedures to be used for addressing survey attrition
- How will the study deal with outcomes with limited variation?
- If you are going to be testing a model, include the model
- Remember to archive it



- Description of the sample to be used in the study
- Key data sources
- Hypotheses to be tested throughout the causal chain
- Specify how variables will be constructed
- Specify the treatment effect equation to be estimated
- What is the plan for how to deal with multiple outcomes and multiple hypothesis testing?
- Procedures to be used for addressing survey attrition
- How will the study deal with outcomes with limited variation?
- If you are going to be testing a model, include the model
- Remember to archive it



- Description of the sample to be used in the study
- Key data sources
- Hypotheses to be tested throughout the causal chain
- Specify how variables will be constructed
- Specify the treatment effect equation to be estimated
- What is the plan for how to deal with multiple outcomes and multiple hypothesis testing?
- Procedures to be used for addressing survey attrition
- Mow will the study deal with outcomes with limited variation?
- If you are going to be testing a model, include the model
- Remember to archive it



- Description of the sample to be used in the study
- Key data sources
- Hypotheses to be tested throughout the causal chain
- Specify how variables will be constructed
- Specify the treatment effect equation to be estimated
- What is the plan for how to deal with multiple outcomes and multiple hypothesis testing?
- Procedures to be used for addressing survey attrition
- 4 How will the study deal with outcomes with limited variation?
- If you are going to be testing a model, include the model
- Remember to archive it



- Description of the sample to be used in the study
- Key data sources
- Hypotheses to be tested throughout the causal chain
- Specify how variables will be constructed
- Specify the treatment effect equation to be estimated
- What is the plan for how to deal with multiple outcomes and multiple hypothesis testing?
- Procedures to be used for addressing survey attrition
- How will the study deal with outcomes with limited variation?
- If you are going to be testing a model, include the model
- Remember to archive it



- Description of the sample to be used in the study
- Key data sources
- Hypotheses to be tested throughout the causal chain
- Specify how variables will be constructed
- Specify the treatment effect equation to be estimated
- What is the plan for how to deal with multiple outcomes and multiple hypothesis testing?
- Procedures to be used for addressing survey attrition
- How will the study deal with outcomes with limited variation?
- If you are going to be testing a model, include the model
- Remember to archive it



```
http://blogs.worldbank.org/impactevaluations/
a-pre-analysis-plan-checklist
```

- Description of the sample to be used in the study
- Key data sources
- 4 Hypotheses to be tested throughout the causal chain
- Specify how variables will be constructed
- Specify the treatment effect equation to be estimated
- What is the plan for how to deal with multiple outcomes and multiple hypothesis testing?
- Procedures to be used for addressing survey attrition
- How will the study deal with outcomes with limited variation?
- If you are going to be testing a model, include the model
- Remember to archive it



- Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article.
- Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.
- Authors must list all variables collected in a study
- Authors must report all experimental conditions, including failed manipulations.
- If observations are eliminated, authors must also report what the statistical results are if those observations are included.
- If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.

- Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article.
- Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.
- Outhors must list all variables collected in a study.
- 4 Authors must report all experimental conditions, including failed manipulations.
- If observations are eliminated, authors must also report what the statistical results are if those observations are included.
- If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.

- Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article.
- Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.
- Authors must list all variables collected in a study.
- 4 Authors must report all experimental conditions, including failed manipulations.
- If observations are eliminated, authors must also report what the statistical results are if those observations are included.
- If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.

- Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article.
- Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.
- Authors must list all variables collected in a study.
- Authors must report all experimental conditions, including failed manipulations.
- If observations are eliminated, authors must also report what the statistical results are if those observations are included.
- If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.

- Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article.
- Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.
- Authors must list all variables collected in a study.
- Authors must report all experimental conditions, including failed manipulations.
- If observations are eliminated, authors must also report what the statistical results are if those observations are included.
- If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.

- Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article.
- Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.
- Authors must list all variables collected in a study.
- Authors must report all experimental conditions, including failed manipulations.
- If observations are eliminated, authors must also report what the statistical results are if those observations are included.
- If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.

- Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article.
- Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.
- Authors must list all variables collected in a study.
- Authors must report all experimental conditions, including failed manipulations.
- If observations are eliminated, authors must also report what the statistical results are if those observations are included.
- If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.

Examples

J-PAL Hypothesis Registry (9)

Conclusion

- Not just for RCTs.
- Spectrum from confirmatory to exploratory research, all has value.
- I'd just like know which research is which.

Conclusion

- Not just for RCTs.
- Spectrum from confirmatory to exploratory research, all has value.
- I'd just like know which research is which.

Conclusion

- Not just for RCTs.
- Spectrum from confirmatory to exploratory research, all has value.
- I'd just like know which research is which.