

Fiabilité de la recherche: Problèmes et pistes de solution



Zacharie Tsala Dimbuene

Juillet 19-20, 2016

Atelier sur la transparence de la recherche

Université de Kinshasa

Kinshasa, République Démocratique du Congo

Part I: Problèmes



Brise-glace...

Prenez quelques minutes (seul ou en groupe)
pour répondre aux questions suivantes:

- 1) Quand vous lisez / entendez quelque chose à propos d'une nouvelle recherche ou publication, quels sont les éléments qui poussent à croire aux résultats? Qu'est-ce qui vous rendrait sceptique?
- 2) Y a-t-il des cas où vous avez essayé de creuser davantage les résultats d'une recherche? Qu'avez-vous fait, et comment cela s'est passé?

Recherche fiable

Les résultats de recherche **fiables** sont extrêmement importants. En effet, les politiques, programmes et produits (tels que les médicaments) qui en résultent peuvent affecter de milliers de vie.

... La question: **Pouvons-nous faire confiance à tous les résultats de recherche?** Quels sont les problèmes qui peuvent pousser à **douter** des résultats d'une recherche? Comment peut-on améliorer ou remédier à cela?

Problèmes dans la recherche:

- Biais de publication
- P-hacking
- Non-divulgation
- Sélection des résultats
- Impossibilité de répliquer
- Manque de transparence

Biais de publication

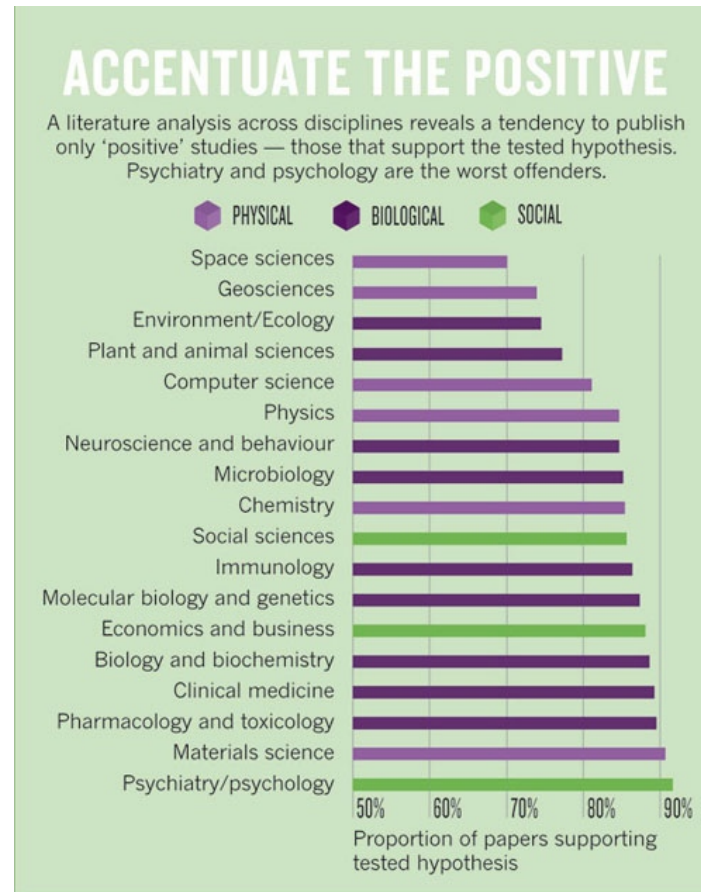
“File drawer problem”



Biais de publication

- Les résultats **statistiquement significatifs** ont plus de chances d'être publiés; ceux qui ne sont pas significatifs sont mis dans les tiroirs.

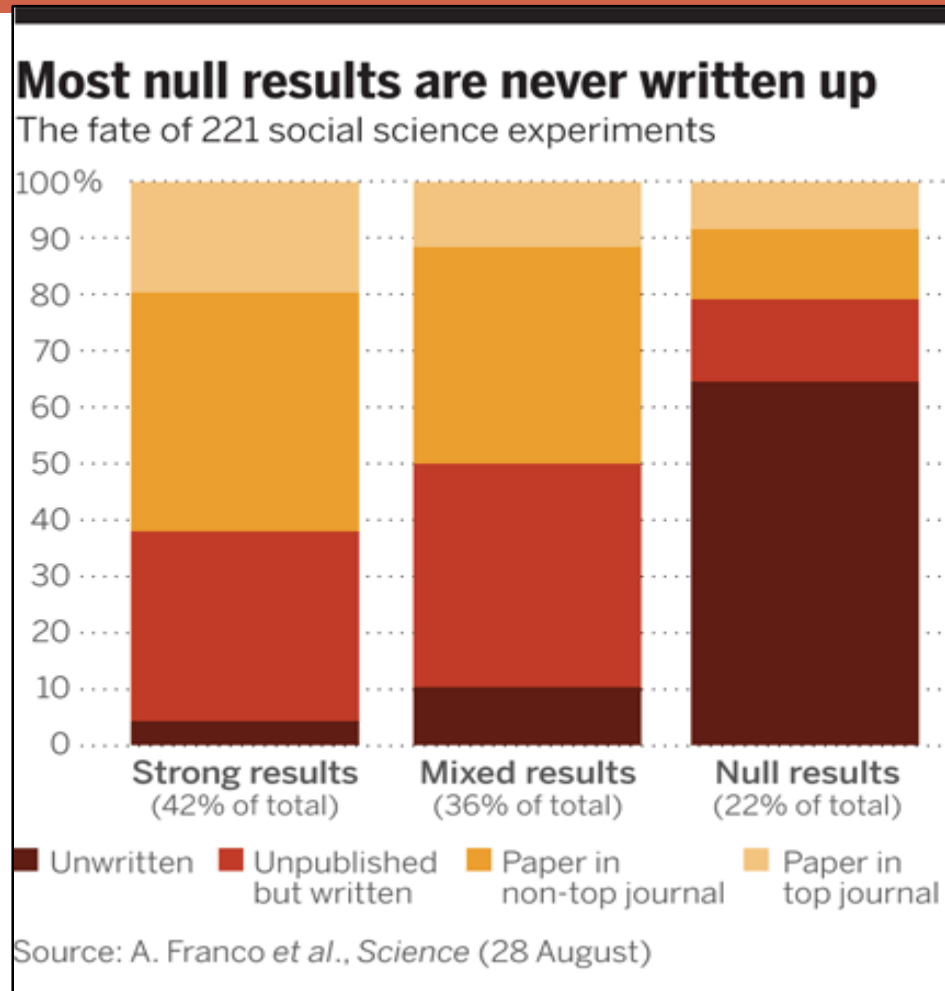
Très peu d'études rapportent les résultats nuls



Biais de publication

- Exemple: Franco et al. ([Science 2014](#)); ils trouvent **que dans une série d'études financées sous NSF à travers TESS (Time-sharing Experiments in the Social Sciences), 60% avec des résultats positifs, et statistiquement significatifs ont été publiés contre seulement 20% avec des résultats non significatifs.**

En science sociales...



Franco et al. ([Science 2014](#))

Pop quiz



- Que répondriez-vous à cette personne?

“Je ne vois aucun problème si les résultats non significatifs ne sont pas rapportés. Nous sommes intéressés à découvrir ce qui fonctionne, et lire les recherches dans lesquelles les résultats ne sont pas significatifs ne nous aident pas mieux!

.... Alors quel est le problème?”

P-Hacking

- Appelé aussi “**data fishing**,” “**data mining**.”
- “Si tu tortures tes données assez longtemps, elles finissent par révéler quelque chose.” -Ronald Coase



"If you don't reveal some insights soon, I'm going to be forced to slice, dice, and drill!"

P-Hacking

- Tester les hypothèses i.e. rechercher les relations entre variables (e.g. scolarisation, résultats des examens).
 - En particulier, un résultat statistiquement significatif à **$p < .05$** est souvent plus intéressant et il a plus de chance d'être publié.
- Ce qui conduit à certaines pressions dans l'**opérationnalisation des variables** (Codage des variables, élimination des valeurs aberrantes, quels sous-groupes à considérer, etc) pour trouver un résultat significatif (**$p < .05$**).

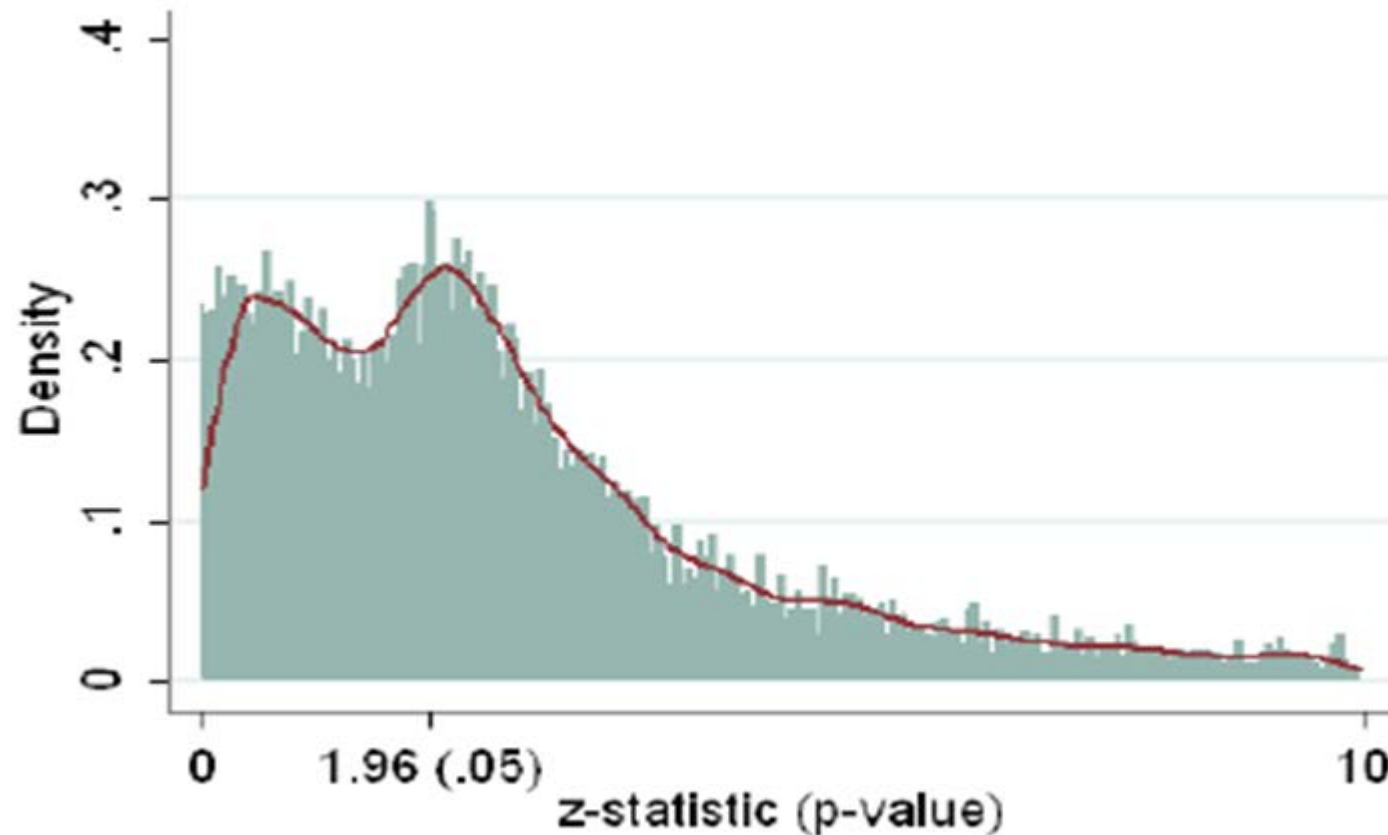
P-Hacking

- “Why Most Research Findings Are False”
Ioannidis ([PLOS 2005](#))

“Plus grande est la flexibilité dans le protocole de recherche, les définitions, variables dépendantes, et les modèles analytiques dans une discipline scientifique, moins les résultats de recherche seront vrais.”

En économie...

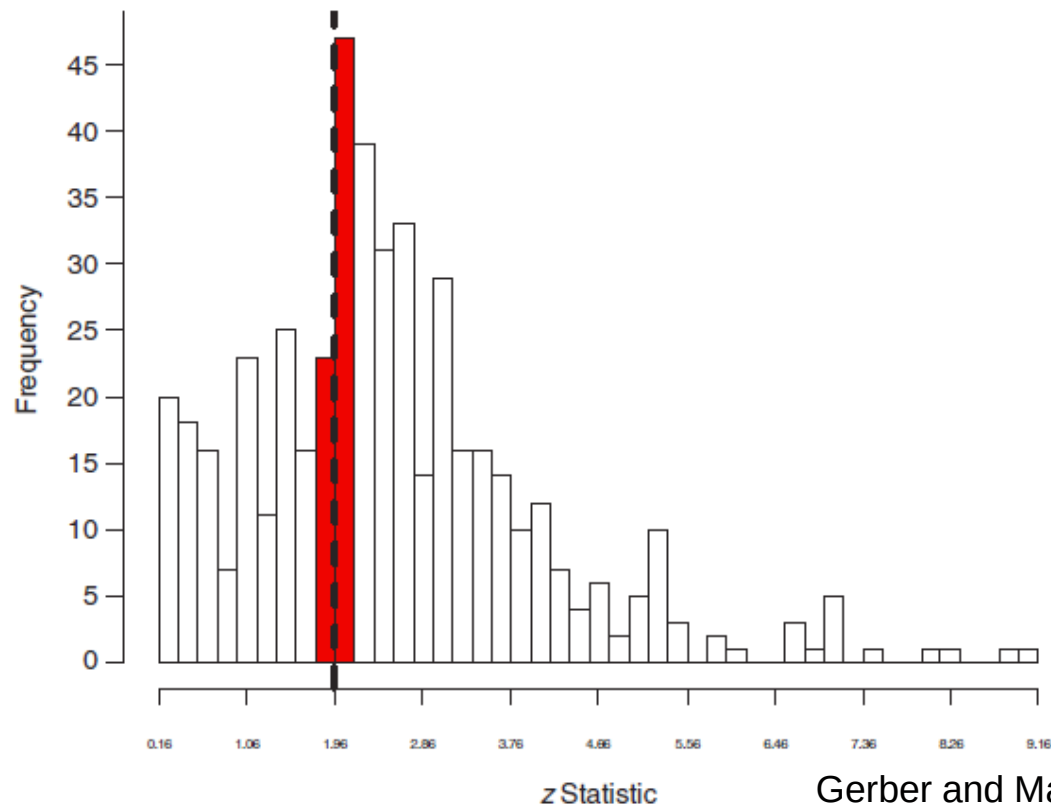
Brodeur et al ([AEJ 2016](#)). Données: 50,000 tests sur des publications dans *AER*, *JPE*, *QJE* (2005-2011)



En sociologie...



Histogram of z Statistics From the *American Sociological Review*, the *American Journal of Sociology*, and *The Sociological Quarterly* (Two-Tailed)



Gerber and Malhotra 2008
doi: 10.1177/0049124108318973

En sciences politiques...



ipa
INNOVATIONS FOR
POVERTY ACTION



BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

Gerber and Malhotra 2008

doi: 10.1561/100.00008024

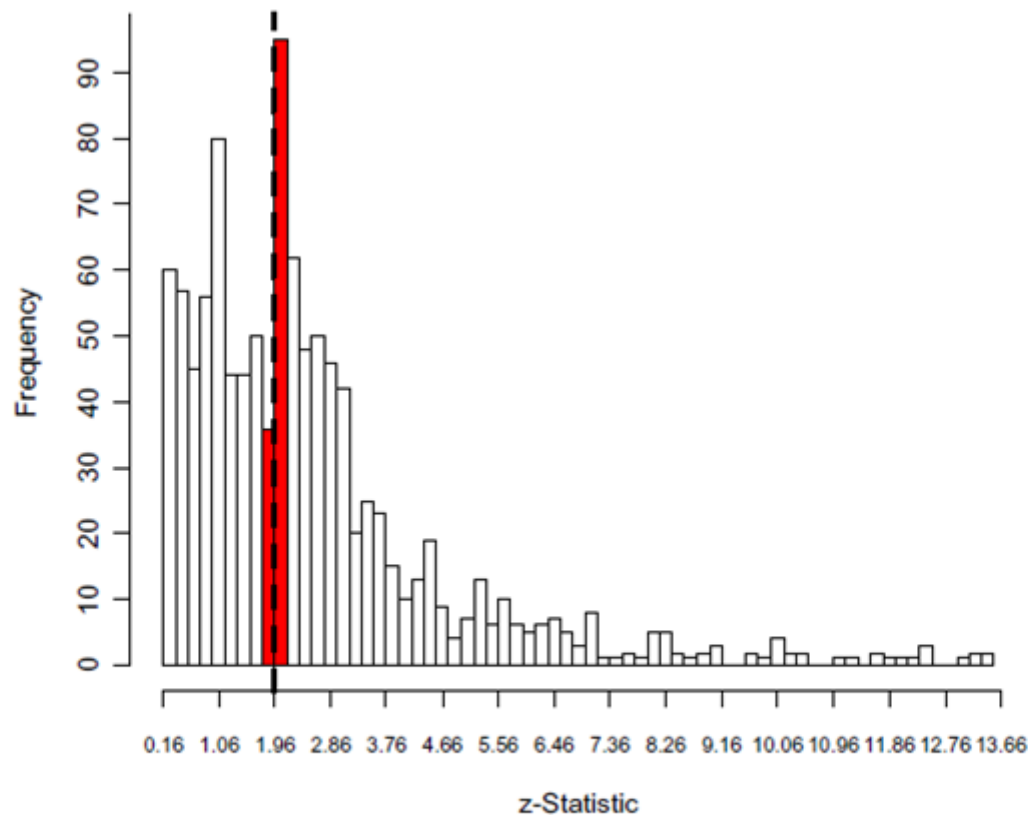
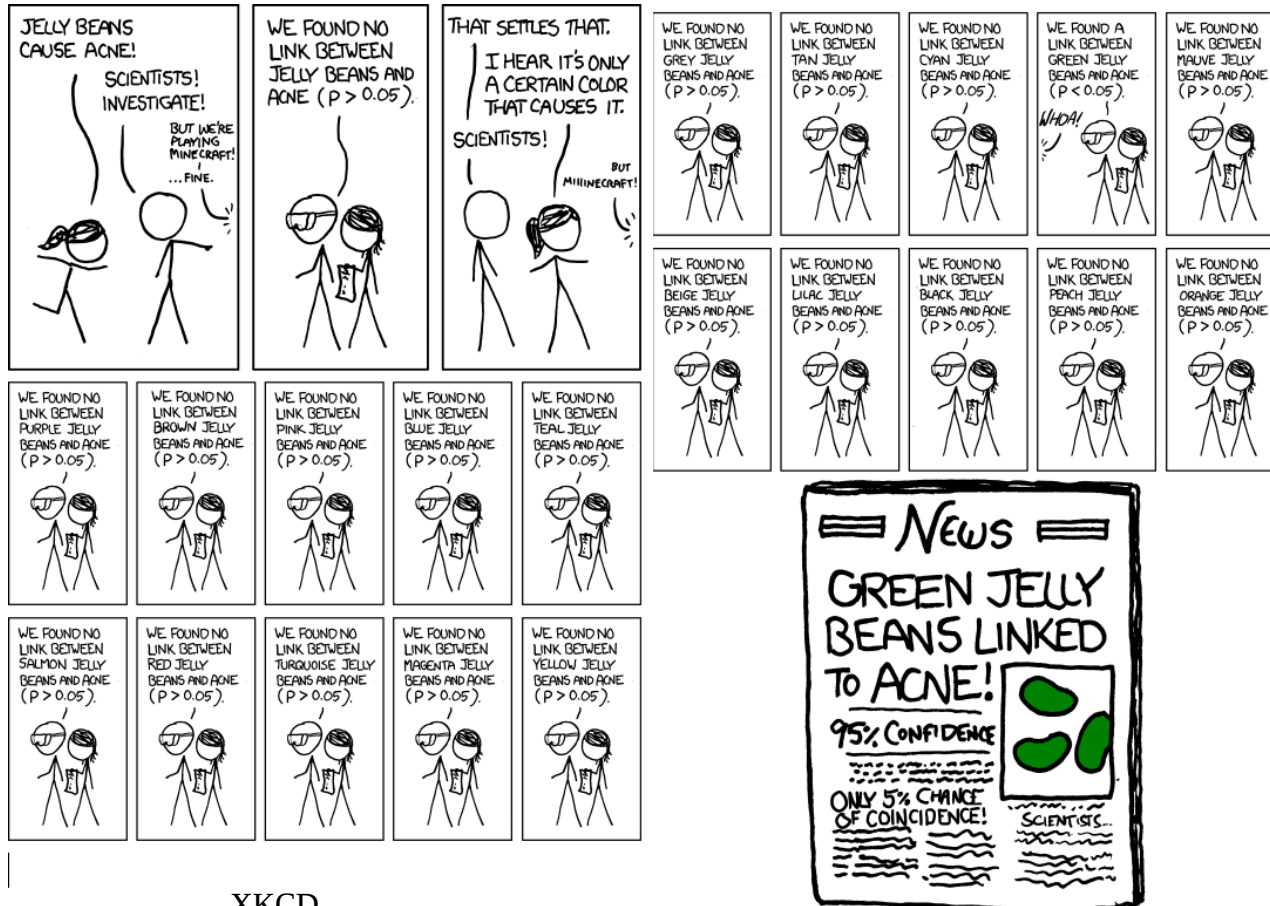


Figure 1(a). Histogram of z -statistics, *APSR* & *AJPS* (Two-Tailed).

Sélection des résultats

- Cherry-picking pour les résultats à présenter

Sélection des résultats



XKCD

<http://xkcd.com/882/>

Sélection des résultats

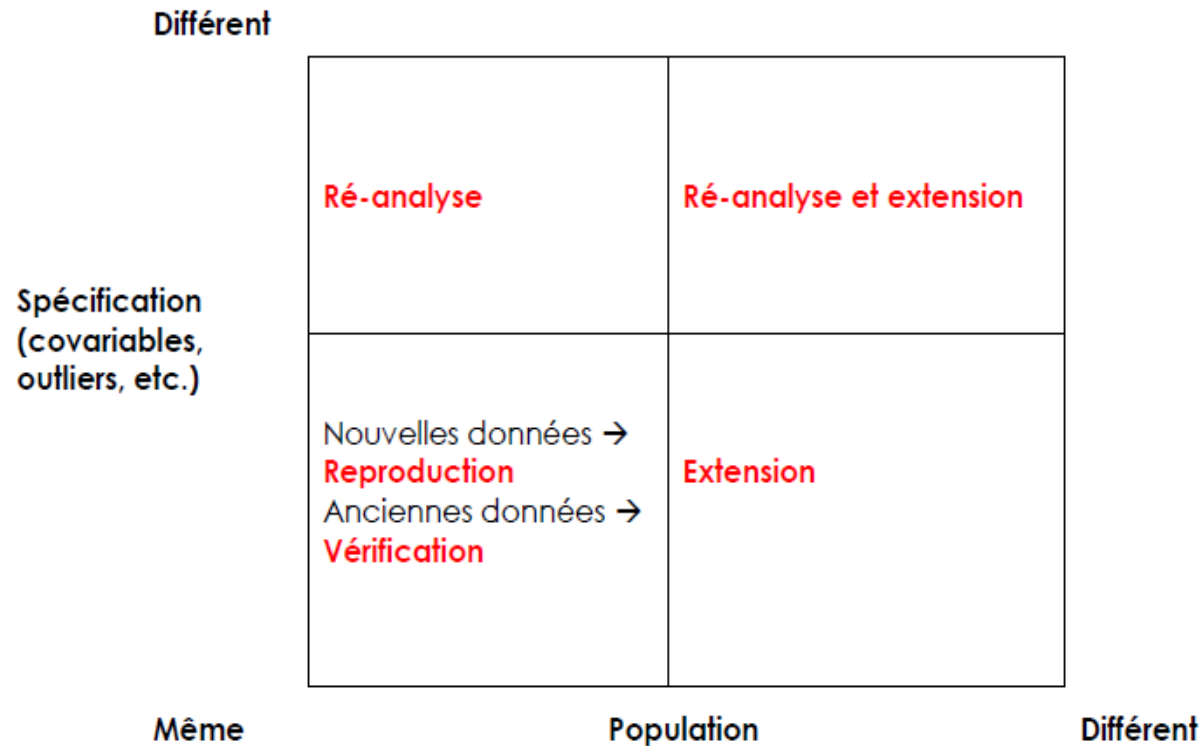
- Malhotra, Franco, Simonovits (2015) trouvent que dans certaines études conduites sous TESS, approximativement 60% des études ne rapportent que peu de variables dépendantes que celles listées dans le questionnaire.
- Si plusieurs “outcomes” ne sont pas mentionnés dans le papier (et ne sont listés nulle part), comment peut-on être confiant que les résultats présentés ne sont pas qu’un bruit, plutôt que de vrais effets?

Réplifications

“Réplication” peut être utilisée pour signifier plusieurs choses. Dans ce contexte, elle signifie:

- 1) **Vérification et re-analyse:** Vérifier que les données originales/code produisent les résultats publiés, et même aller au-delà pour vérifier la robustesse des résultats.
- 2) **Reproduction:** Tester si les résultats tiennent toujours quand l'étude est conduite de façon similaire.
- 3) **Extension:** Démontrer si les résultats tiennent lorsque l'étude est conduite dans un autre contexte, sous différentes conditions, etc., en vue de tester la validité externe.

Réplifications



Terminology from [Michael Clemens 2014](#)

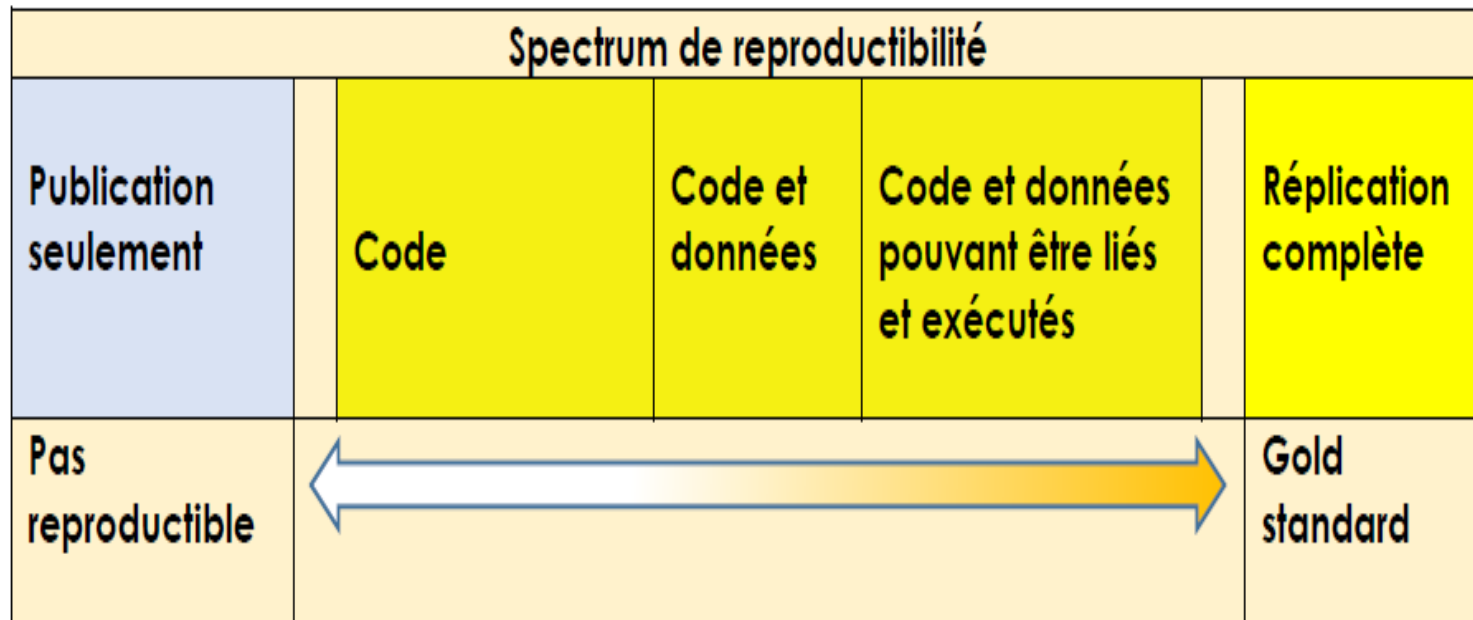
Quelques exemples de réplifications:

- Reinhart-Rogoff (2013)
 - Erreurs trouvées sur une feuille de calcul par un étudiant aux études supérieures ([Herndon et al 2013](#)).
- Deworming debate (2015)
 - [Miguel/Kremer](#) (2004) and [Aiken/Davey et al.](#) (2015)
 - Grand débat au sein de l'économie du développement et l'épidémiologie
- [Reproducibility project](#) in psychology (2015)
 - ~40% d'études reproduites avec succès.
 - Mais grand débat la semaine passée! Que signifie qu'une réplification "a échoué"?
- [Begley et al.](#) (2012):
 - Tentative de répliquer des études de laboratoire sur le cancer pré-cliniques chez (6 sur 53 études ont été répliquées).

Motivation pour les répliques

- Les revues sont souvent sceptiques de publier des travaux qui ne sont pas “**originaux**”; donc les répliques sont considérées comme “**non originales**.”
- Pourtant poster les résultats des répliques sur un website (e.g. **Political Science Replication Initiative; Psych file drawer**) n’est pas idéal non plus: certains pensent que même les répliques devraient être évaluées par un comité des pairs.

Transparence



Peng 2011, Science 334(6060) pp. 1226-1227

Transparence

- Partage des données, code et enquêtes (avec une documentation claire) peut permettre aux autres de vérifier le travail de quelqu'un.

Transparence



- Pourtant peu de chercheurs partagent les données:
 - [Alsheikh-Ali et al. 2011](#): Revue de 10 premiers papiers de 2009 publiés dans les 50 top journals classés par facteur d'impact. Parmi 500 papiers, 351 devaient publier aussi les données: 59% n'ont pas adhéré à ces politiques, en ne publiant pas les données. En somme, seulement 47 papiers (9%) ont mis en ligne les données brutes.
- Les agences subventionnaires (e.g. Fonds de gouvernement pour la recherche telles que NSF, NIH, ERSC au UK, fondations privées) sont entrain d'adopter des politiques de partage des données. Par contre, elles n'obligent pas nécessairement.

Transparence



- Relativement peu des revues ont des politiques de partage des données:
 - En 2013, 18 sur 120 revues de science politiques avaient une telle politique ([Gherghinaa and Katsanidoua 2013](#))
 - Une autre étude a trouvé que 29 sur 141 revues en économie avaient cette politique ([Vlaeminck 2013](#)).
- Une politique obligatoire dans ce sens pousserait les chercheurs plus enclins à partager les données:
 - Dans les journaux qui imposent la mise à disposition des chercheurs des données de recherche, le taux de partage est 1000x plus élevé que dans les journaux qui n'ont pas une telle politique. ([Vines 2014](#))

Discussion

Discussion en groupes de 5-6:

- Quels changements dans la façon dont les recherches sont menées pourraient **mieux** améliorer leur fiabilité? Pourquoi?

Part II: Solutions



Méta-recherche



Comme discuté, les problèmes incluent:

- Biais de publication
- Data-mining
- Sélection des résultats à reporter
- Manque de transparence et impossibilité de réplication

Comment améliorer?



Transparence dans la recherche!

1. Enregistrement des études et plans de pré-analyse
2. “Résultats-neutres publiés”
3. Réplications
4. Publier / partager tous les résultats de l'étude
5. Partager les données: données, codes, enquêtes, fichiers “lisez-moi”

Enregistrer les études

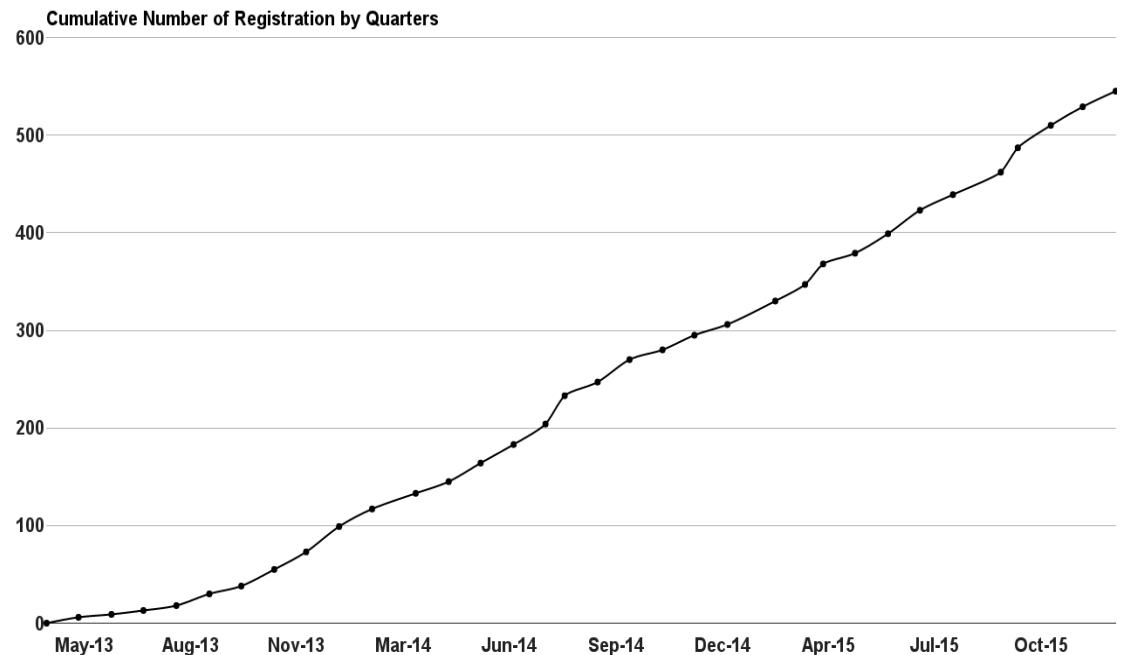


- Créer un enregistrement public de l'étude avec des informations de base (e.g., intervention, outcomes, location, dates).
- Toutes les études menées par IPA et J-PAL sont encouragées de pré-enregistrer sur le site (www.socialscienceregistry.org).
 - Il est recommandé de pré-enregistrer avant de commencer l'intervention.

Ce qu'est le registre "AEA" ...



- Un website pour enregistrer les essais randomisés et contrôlés (ERC) en sciences sociales, établi par **American Economics Association** (AEA)
- Établi depuis 2013
- ~580 ERC enregistrés.



C'est quoi le registre AEA?





AEA RCT Registry

The American Economic Association's registry for randomized controlled trials

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

Advanced Search [SEARCH](#)

ACA RCT Registry currently lists 503 studies with locations in 80 countries. [REGISTER A TRIAL >](#)

MOST RECENTLY REGISTERED TRIALS

Identity and Electoral Processes: Evidence from Bihar

LAST REGISTERED ON MARCH 01, 2016 

In 2014, the state of Bihar randomly assigned individuals to the teams of election officers who manage polling stations on election day. This ongoing study represents an effort to understand how the composition of the teams of polling station officials may impact voting outcomes as well as behaviors and perceptions of voters and election officers.

[VIEW TRIAL >](#)

Preventing Excess Female School Drop Out in Mozambique: Conditional Transfers and the Respective Role of Parent and Child in Schooling Decisions

LAST REGISTERED ON FEBRUARY 29, 2016 

In order to shed light on the respective role of parents and children in making schooling decisions, our experiment will assign primary schools in Mozambique's Manica province randomly across four groups, comprising three treatment groups and one

Autres registres:

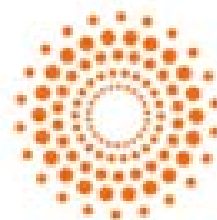
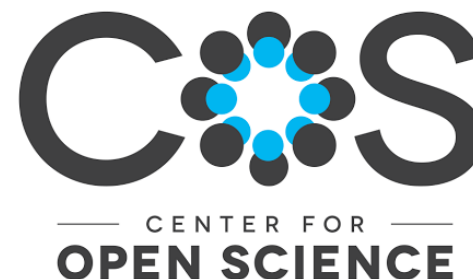


ClinicalTrials.gov

A service of the U.S. National Institutes of Health



RIDIE
Registry for International
Development Impact Evaluations



egap EVIDENCE
IN GOVERNANCE
AND POLITICS



Récompenses pour enregistrement



CENTER FOR OPEN SCIENCE

About us ▾

Services ▾

Get Involved ▾

Communities

News

Donate

The \$1,000,000 Preregistration Challenge

The Big Picture

The Challenge

How to Earn the Prize

Eligibility Criteria

FAQ

Eligible Journals

Review Process

Preregistration increases the credibility of hypothesis testing by confirming in advance what will be analyzed and reported. For the Preregistration Challenge, one thousand researchers will win \$1,000 each for publishing results of preregistered research.

Share [this handout](#) for a brief overview and links to more information, and [begin your preregistration today!](#)

Begin a
Preregistration



Pourquoi enregistrer son étude?

- Combat le biais de publication en produisant un numéro de l'étude au grand public.
- Pré-enregistrement est une **exigence de longue date en médecine** (e.g., pour approbation auprès de FDA, publication dans les revues médicales) et devient de plus en plus important en sciences sociales.
 - [Clinicaltrials.gov](https://clinicaltrials.gov) (185K+ études enregistrées).



Comment et où enregistrer?

- Comment?
 - Créer un compte sur socialscienceregistry.org.
 - 18 champs obligatoires: information de base telle que le titre de l'essai, pays, dates de début et fin de l'intervention, outcomes, etc.
 - Personnel du projet (e.g. gestionnaires de la recherche) peuvent enregistrer un essai pour le chercheur principal (CP), mais le CP doit aller sur le site pour approuver et soumettre l'enregistrement.
 - Plus de détails sur “comment s’enregistrer” et une video sur Box [\[link\]](#) / Sharepoint.

Plans de Pré-analyse

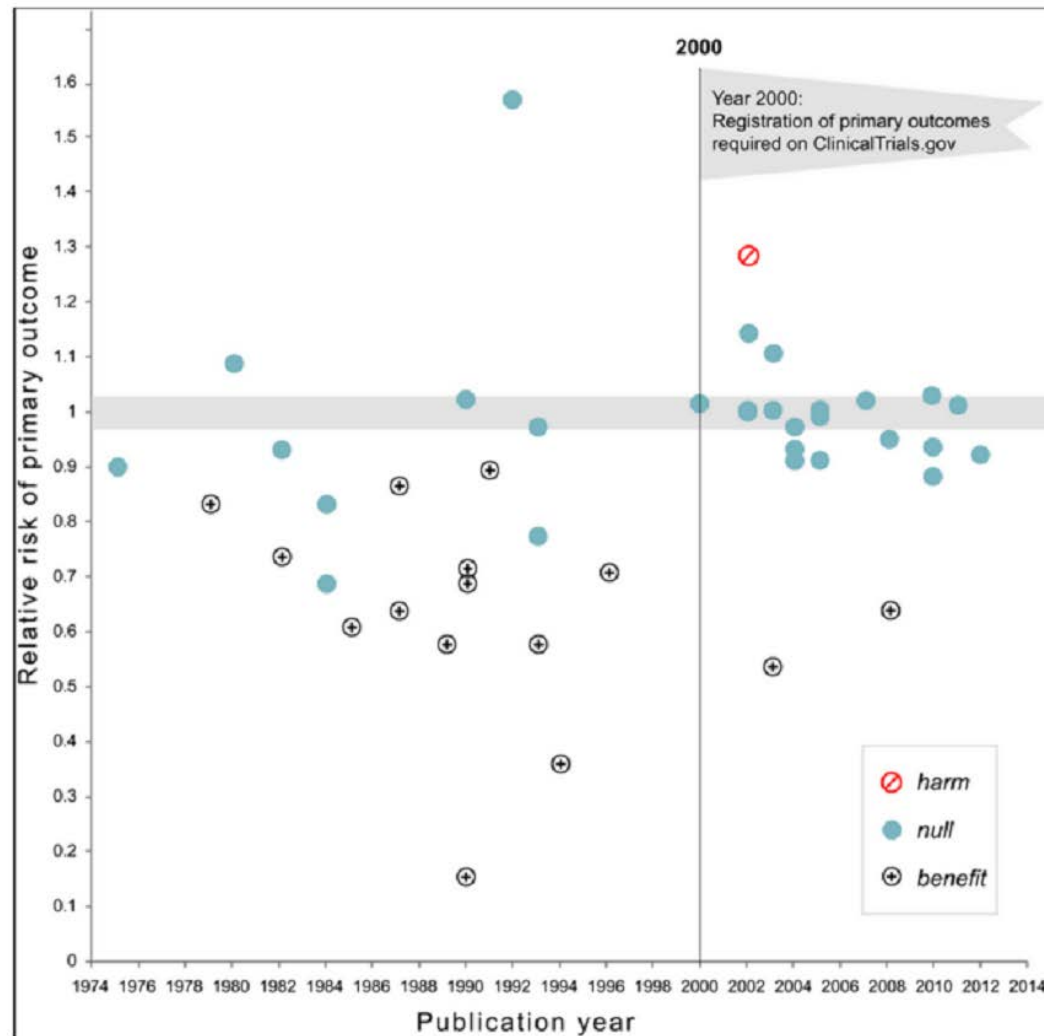
- Les plans de pré-analyse sont des documents plus détaillés sur les hypothèses, outcomes, et les analyses qui sont planifiées.
 - L'objectif est de combattre le “data-mining” en contraignant le chercheur à **“s'accrocher sur ce qu'il avait en tête au début de l'étude”**
- Le registre AEA n'exige pas les plans de pré-analyse (près de 1/6 d'études enregistrées ont de plan de pré-analyse
 - ✓ Vous pouvez faire une recherche des “plans de pré-analyse” si intéressé).

Plans de pré-analyse

Pourquoi un plan de pré-analyse?

Nous avons identifié tous les large essais randomisés et contrôlés financés par NHLBI entre 1970 et 2012 évaluant les médicaments ou suppléments alimentaires pour le traitement ou la prévention des maladies cardiovasculaires. 17 sur 30 études (57%) publiées avant 2000 ont montré un bénéfice significatif de l'intervention sur le "outcome" contre seulement 2 sur 25 (8%) essais publiés après 2000. ([Kaplan and Irvin, 2015, PLOS](#)).

Plans de pré-analyse



Partager tous les résultats (y compris les résultats non significatifs!)

- Le biais de publication conduit à un **étalement** vers les résultats positifs (excitants).
 - Problème: comment disséminer les résultats non significatifs, si les revues ont tendance à accepter les résultats positifs à un taux plus élevé?
- Une solution: reporter tous les résultats dans le registre public.
 - Par exemple: la FDA aux É-U exige que tous les résultats soient publiés sur clinicaltrials.gov dans l'année qui suit la fin de l'essai. Pour les sciences sociales, les résultats pourraient être publiés là où l'étude a été enregistrée (e.g. registre AEA).
- Alltrials est un groupe qui bat campagne pour pousser tous les essais cliniques à publier les résultats.

Publier les résultats-neutres

- Soumettre un protocole d'étude avec tous les détails sur la méthodologie de l'étude et les analyses planifiées, avant même de mener l'étude. L'article est accepté sur base de protocole, avant que les données soient collectées.
- “Puisque l'étude est acceptée en avance, les motivations des chercheurs changent en produisant l'histoire **correcte** plutôt que l'histoire la plus belle.” –Chris Chambers, editor of Cortex (OSF [citation](#))

Publier les résultats-neutres...

20 revues ont adopté le modèle des “rapports enregistrés” ([list](#)). Exemples:

- AIMS Neuroscience
- Attention, Perception & Psychophysics
- Comparative Political Studies
- Comprehensive Results in Social Psychology
- Cortex
- Drug and Alcohol Dependence
- eLife
- Experimental Psychology
- Frontiers in Cognition
- Journal of Business and Psychology
- Nutrition and Food Science Journal
- Perspectives on Psychological Science
- Social Psychology
- Working, Aging, and Retirement

Réplifications

Aussi valable qu'elles soient, les réplifications sont relativement rares.

Que peut faire la communauté scientifique pour que le nombre de réplifications augmentent, soient faites et partagées?

Quelques initiatives existent:

- Programme de replication 3ie'
- Projet de reproductibilité en psychologie et biologie du cancer
- Réplication wiki in économie

Partage des données (Data-sharing)



- Qu'est-ce?
 - Rendre les données, les syntaxes ou code, et autres matériels de l'étude **publiquement disponible**.
- "An article...in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete set of instructions [and data] which generated the figures."
 - Dr. David Donoho, Stanford stats professor

Partage des données: Pourquoi?



- Faciliter la replication des résultats publiés
- Re-utilisation des données pour des études futures et la méta-analyse
- Promouvoir une meilleure qualité des données/syntaxes-codes/métadonnées



Le mouvement de “partage des données”

- Bailleurs de fonds:
 - Holdren memo, Février 2013
 - Fondations – Gates, Wellcome Trust (et 15 autres ont signé un communiqué conjoint), et d’autres encore
- Revue et associations
 - Top revues en économie e.g. AER, Econometrica, AEJ
 - American Political Science Association – nouvelles directives éthiques sur le partage des données
- Open science groups font un plaidoyer sur open data
 - Berkeley Initiative for Transparency in the Social Sciences
 - Center for Open Science
 - Meta-research institute at Stanford (Ioannidis)
 - YODA at Yale (clinical trials data-sharing portal)



Transparence de la recherche à l'IPA

- IPA a adopté de nouvelles directives et une politique de publication des données trois ans après la collecte.
- Répositoire des données pour les ERCs (incluant IPA/J-PAL et autres organisations partenaires).



ipa
INNOVATIONS FOR
POVERTY ACTION



Données des ERCs en sciences sociales



About

Guides ▾

Support

Stephanie Wykstra ▾

Randomized Controlled Trials in the Social Sciences Dataverse (Innovations for Poverty Action)

Harvard Dataverse > Randomized Controlled Trials in the Social Sciences Dataverse



Edit ▾

Data from RCTs in the social sciences, including data from Innovations for Poverty Action and J-PAL studies. Data, statistical code underlying published research results, surveys and documentation are included. Contact for questions/comments: researchsupport@poverty-action.org.

Search this dataverse...

Find

Advanced Search

+ Add Data ▾

☒ Datasets (1)

☒ Datasets (48)

☐ Files (1,110)

Publication Status

Published (48)

Draft (1)

Publication Date

2014 (14)

2015 (13)

2007 (7)

2013 (4)

2009 (3)

More...

Subject

Social Sciences (25)

Medicine, Health and Life Sciences

(1)

Author Name

Esther Duflo (12)

Karlan, Dean (11)

Duflo, Esther (7)

Abhijit Banerjee (5)

1 to 10 of 49 Results

Sort ▾

«

< Previous

1

2

3

4

5

Next >

»

Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State

Aug 17, 2015



Blattman, Chris; Annan, Jeannie, 2015, "Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State", <http://dx.doi.org/10.7910/DVN/11R0LX>, Harvard Dataverse, V1 [UNF:6:kvfhTCYS34Zqc/5hMQ=]

States and aid agencies use employment programs to rehabilitate high-risk men in the belief that peaceful work opportunities will deter them from crime and violence. Rigorous evidence is rare. We experimentally evaluate a program of agricultural training, capital inputs, and coun...

Microcredit Impacts: Evidence from a Randomized Microcredit Program Placement Experiment by Compartamos Banco

Jul 13, 2015



Angelucci, Manuela; Karlan, Dean; Zinman, Jonathan, 2014, "Microcredit Impacts: Evidence from a Randomized Microcredit Program Placement Experiment by Compartamos Banco", <http://dx.doi.org/10.7910/DVN/27850>, Harvard Dataverse, V3

We use a clustered randomized trial, and over 16,000 household surveys, to estimate impacts at the community level from a group lending expansion at 110 percent APR by the largest microlender in Mexico. We find no evidence of transformative impacts on 37 outcomes (although some e...

Finding Missing Markets (and a disturbing epilogue): Evidence from an Export Crop Adoption and Marketing Intervention in Kenya

Jul 12, 2015



Ashraf, Nava; Giné, Xavier; Karlan, Dean, 2015, "Finding Missing Markets (and a disturbing epilogue): Evidence from an Export Crop Adoption and Marketing Intervention in Kenya", <http://dx.doi.org/10.7910/DVN/PES3RD>, Harvard Dataverse, V1

Farmers may grow crops for local consumption despite more profitable export options. DrumNet, a Kenyan NGO that helps small farmers adopt and market export crops, conducted a randomized trial to evaluate its impact. DrumNet services increased production of export crops and lowered...

Ghana Youth Survey (GYS)

Jul 2, 2015



Duflo, Esther; Dupas, Pascaline; Kremer, Michael, 2015, "Ghana Youth Survey (GYS)", <http://dx.doi.org/10.7910/DVN/CHNG58>, Harvard

Données des ERCs en sciences sociales

Randomized Controlled Trials in the Social Sciences Dataverse (Innovations for Poverty Action)

Harvard Dataverse > Randomized Controlled Trials in the Social Sciences Dataverse >
 Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State

Metrics 1 Download

Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State

Blattman, Chris; Annan, Jeannie, 2015, "Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State", <http://dx.doi.org/10.7910/DVN/11ROLX>, Harvard Dataverse, V1 [UNF6:kvfhtTCY8t34IZqcShIMQ==]

If you use these data, please add this citation to your scholarly resources. Learn about Data Citation Standards.

Description
 States and aid agencies use employment programs to rehabilitate high-risk men in the belief that peaceful work opportunities will deter them from crime and violence. Rigorous evidence is rare. We experimentally evaluate a program of agricultural training, capital inputs, and counseling for Liberian ex-fighters who were illegally mining or occupying rubber plantations. 14 months after the program ended, men who accepted the program offer increased their farm employment and profits, and shifted work hours away from illicit activities. Men also reduced interest in mercenary work in a nearby war. Finally, some men did not receive their capital inputs but expected a future cash transfer instead, and they reduced illicit and mercenary activities most of all. The evidence suggests that illicit and mercenary labor supply responds to small changes in returns to peaceful work, especially future and ongoing incentives. But the impacts of training alone, without capital, appear to be low.

Subject
 Social Sciences

Keyword
 Conflict, Crime, Violence, Illicit activity, Agriculture, Training program, Employment, Soldier, Mercenary, Cash transfer

Files Metadata Terms Versions

7 Files

ado_files.zip
 ZIP Archive - 157.1 KB - Aug 17, 2015 - 0 Downloads
 MD5: 5aead39c1074174b48823b79b94485
 User-written commands needed to run the code.
[ADO files](#) [Download](#)

analysis_final.tab
 Tabular Data - 10.8 MB - Aug 17, 2015 - 1 Download
 Original File MD5: d72d61c3d95f63ae2f5b6ed3d3d2f
 3105 Variables, 1274 Observations - UNF: 6:kvfhtTCY8t34IZqcShIMQ==
 Data dataset to produce all tables and figures. Contains all data collected from the study.
[Data](#) [Explore](#) [Download](#)

Baseline Survey.pdf
 Adobe PDF - 258.8 KB - Aug 17, 2015 - 0 Downloads
 MD5: ca8e69a23d3118ac3c39f7aada7f8d3
 Baseline survey.
[Surveys](#) [Download](#)

Lignes directrices chez IPA



Guidelines for data publication

February 26, 2015

Introduction

Journals, funders and research organizations are increasingly moving towards greater research transparency, including sharing materials such as data and code. Sharing research materials beyond published articles is valuable both because it allows others to examine the fuller body of evidence for reported results, and also because it facilitates fuller re-use of collected data.

Innovations for Poverty Action has joined the movement towards greater transparency, by adopting a [data-sharing requirement](#) and creating a public [repository](#) for sharing data. We invite researchers to submit their materials to our data repository team, which provides assistance in preparing the data and code and uploading it for publication.

This guide is for IPA affiliates, and includes our guidelines for data-sharing as well as an explanation of IPA's data curation process. For question and comments, please contact Stephanie Wykstra at researchsupport@poverty-action.org.

Contents

- [Why share data and code?](#)
- [IPA's requirement](#)

Que peut-on partager?

1. Bases de données

- Recommandé: base de données de l'étude nettoyée (anonymisée!)
- Minimale: base de données de la “publication” – code/données pour la publication.
- Idéal: début pour la reproductibilité (plus sur ceci dans quelques minutes!)

2. Fichiers “Lisez-moi” qui expliquent les relations entre les données et syntaxes ou code, ainsi que toute documentation sur les données.

3. Enquêtes

4. Les métadonnées de l'étude

Où partager les données?

Répositoires des données:

- Dataverse (utilisé par IPA et J-PAL)
 - Vous pouvez créer votre propre répositoire des données en utilisant dataverse (sans frais!) e.g. Dataverse de Ted Miguel.
- Figshare
- Dryad
- ICPSR (n'est pas gratuit pour mettre les données, mais les services de curation sont offerts).
- D'autres sources sur spécifiques types des données existent, voir [Re3data.org](https://re3data.org)

Étapes avant de partager les données/code:

1. Vérifier la convivialité des données e.g., les “labels/value codes” des variables existent
2. S’assurer que les données ont été anonymisées:
 - Identificateurs directs (noms, IDs, etc) sont supprimés.
 - Identificateurs indirects tels que les informations géospatiales sont supprimés.
 - Vérifier si certaines informations peuvent être combinées pour identifier les participants.
3. Réplication de l’étude: vérifier si les données/codes produisent les résultats publiés.
4. Vérifier que les fichiers “Lisez-moi” fournissent de bonnes instructions pour la réplication.
5. S’assurer que variables/fichiers/métadonnées de l’étude sont partagées.

Une “étude complètement reproductible”?

- L'idéal: partager “début—fin” données/code pour permettre la réplication en partant des données brutes aux tableaux finaux.
 - Inclure toutes les syntaxes (construction des variables et nettoyage des données).
- Élément important: partager *quelques* données/codes qui sous-tendent les résultats publiés ... c'est mieux que ne rien partager!

Commencer tôt la préparation des données/codes



Préparer les données/code tôt dans une étude est crucial, il y a en effet de sérieuses limitations pour ce qui pourrait être partagé et compris plus tard dans l'étude, si les bonnes pratiques ne sont pas suivies.

Préparer les données/code

e.g.,

- Nommer et étiquetter les variables
- Commenter les syntaxes
- Master do files
- En-tête/bas de page
- Fichiers “Lisez-Moi”

...Plus de détails dans IPA's [best practice manual](#)
pour gérer les données et syntaxes.

Outils pour la réproductibilité des études

- Contrôle des versions pour les codes en utilisant **git** (plus tard)
- Documents dynamiques en utilisant **Markdoc** (Stata) ou **Rmarkdown** (R) pour générer des tableaux directement dans les articles à partir des syntaxes (demain)
- **Open science framework** pour les travaux de collaboration (demain)

Incitations à partager les données?

Récompenses pour “open” pratiques

March 25, 2014

This Week in *Psychological Science* (TWiPS)

The links below take you to the journal via the APS website. If not already logged in, you will be redirected to log-in using your last name (News) and Member ID (16241).

[Inferring Character From Faces: A Developmental Study](#)
 Emily J. Cogsdill, Alexander T. Todorov, Elizabeth S. Spelke, and Mahzarin R. Banaji

Trustworthiness Dominance Competence

Adults automatically and intuitively make character assessments about people from looking at their faces. Does this ability emerge early in life, or is it something that gradually develops across childhood? Adults and children (ages 3 to 10) were shown pairs of faces with features generated to portray high or low levels of trustworthiness, dominance, or competence. The participants were asked to indicate which of the two faces possessed the respective trait. Children as young as 3 to 4 made reliably consistent inferences about character based on facial characteristics. By age 7, their inferences showed a level of consensus matching that seen in adults, suggesting that the ability to make face-to-trait inferences arises early in life and does not require a prolonged period of social experience to reach an adult level.

[Come listen to Mahzarin R. Banaji give the Fred Kavli Keynote Address "Group Love: How the Mind Creates 'Us' and 'Them'" at the 28th APS Annual Convention in San Francisco, CA, USA.](#)

[The Emergence of Us and Them in 80 Lines of Code: Modeling Group Genesis in Homogeneous Populations](#)
 Kurt Gray, David G. Rand, Eyal Ert, Kevin Lewis, Steve Hershman, and Michael I. Norton

Groups often form based on some identifying dimension, such as race, language, or

Open Data Open Materials Preregistered

Click here for more information

aps 2014 May 22-25 San Francisco



<https://osf.io/tvyxz>

Incitations à partager les données?

Citations des données et aussi un changement dans les pratiques de promotion dans les universités, basé sur non seulement les publications, mais aussi pour récompenser la transparence dans la recherche.

Incitations à partager les données?

Transparence et clarté des lignes directrices dans le processus ([Science](#), 2015)

Summary of the eight standards and three levels of the TOP guidelines				
Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard.				
	LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
Citation standards	Journal encourages citation of data, code, and materials—or says nothing.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used, consistent with journal's author guidelines.	Article is not published until appropriate citation for data and materials is provided that follows journal's author guidelines.
Data transparency	Journal encourages data sharing—or says nothing.	Article states whether data are available and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Analytic methods (code) transparency	Journal encourages code sharing—or says nothing.	Article states whether code is available and, if so, where to access them.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Research materials transparency	Journal encourages materials sharing—or says nothing.	Article states whether materials are available and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Design and analysis transparency	Journal encourages design and analysis transparency or says nothing.	Journal articulates design transparency standards.	Journal requires adherence to design transparency standards for review and publication.	Journal requires and enforces adherence to design transparency standards for review and publication.
Preregistration of studies	Journal says nothing.	Journal encourages preregistration of studies and provides link in article to preregistration if it exists.	Journal encourages preregistration of studies and provides link in article and certification of meeting preregistration badge requirements.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Preregistration of analysis plans	Journal says nothing.	Journal encourages preanalysis plans and provides link in article to registered analysis plan if it exists.	Journal encourages preanalysis plans and provides link in article and certification of meeting registered analysis plan badge requirements.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies—or says nothing.	Journal encourages submission of replication studies.	Journal encourages submission of replication studies and conducts blind review of results.	Journal uses Registered Reports as a submission option for replication studies with peer review before observing the study outcomes.

Discussion

Nous avons parlé de moyens d'améliorer la fiabilité de la recherche (partage des données/code, réplication des travaux, pré-enregistrement des études et présenter tous les résultats):

- Selon vous, quels sont les obstacles principaux qui empêchent les chercheurs à aller dans ce sens?
- Selon votre opinion, quels sont les moyens les plus promoteurs pour améliorer la fiabilité de la recherche? Quels sont ceux moins prometteurs?

-
- Merci
 - Thank you
 - Assante sana