

Adventures in Open and Assumption-Free Research

Charlie Rahal

LCDS and Nuffield College, University of Oxford

A talk to the REAL Demand Unit, 17th December 2024



Beginning to think about Open Science

- Lets randomly allocate ten numbers as evenly as possible!

Beginning to think about Open Science

- Lets randomly allocate ten numbers as evenly as possible!
- Each number pertains to a question which you are going to be asked!

Beginning to think about Open Science

- Lets randomly allocate ten numbers as evenly as possible!
- Each number pertains to a question which you are going to be asked!
- There are no right or wrong answers here!

Beginning to think about Open Science

- Lets randomly allocate ten numbers as evenly as possible!
- Each number pertains to a question which you are going to be asked!
- There are no right or wrong answers here!
- The idea is to help us start thinking about 'Open Science'.

1. "What's one piece of scientific knowledge you think everyone should have access to?"

Thought Experiments

1. "What's one piece of scientific knowledge you think everyone should have access to?"
2. "What's one scientific myth or misconception you wish more people understood correctly?"

Thought Experiments (Cont.)

4. 'What scientific discovery do you think would change the world if everyone knew about it?'

Thought Experiments (Cont.)

4. 'What scientific discovery do you think would change the world if everyone knew about it?'
5. 'What's a research area you think should always be open and accessible to the public?'

Thought Experiments (Cont.)

4. 'What scientific discovery do you think would change the world if everyone knew about it?'
5. 'What's a research area you think should always be open and accessible to the public?'
6. 'If you could design a science experiment that the whole world could contribute to, what would it be?'

Thought Experiments (Cont.)

4. 'What scientific discovery do you think would change the world if everyone knew about it?'
5. 'What's a research area you think should always be open and accessible to the public?'
6. 'If you could design a science experiment that the whole world could contribute to, what would it be?'
7. 'What's one thing scientists can do to make their research more accessible to non-scientists?'

Thought Experiments (Cont.)

8. 'What's one everyday thing you wish more people understood the science behind?'

Thought Experiments (Cont.)

8. 'What's one everyday thing you wish more people understood the science behind?'
9. 'How can we encourage more people to participate in or contribute to open science?'

Thought Experiments (Cont.)

8. 'What's one everyday thing you wish more people understood the science behind?'
9. 'How can we encourage more people to participate in or contribute to open science?'
10. 'If you could crowd-source one big research project, what would you ask people to help with?'

In the spirit of ‘Open Science’:

crahal.github.io/files/real_open_science.pdf

and *all* of the research you see here today is at:

github.com/crahal

Lots of examples: hopefully something for everyone!

Everything here is made with free, and ‘Open’ tools:

- Linux (operating system)
- Python (data wrangling and statistical software)
- $\text{\LaTeX}$ (typesetting)

Everything here is made with free, and ‘Open’ tools:

- Linux (operating system)
- Python (data wrangling and statistical software)
- $\text{\LaTeX}$ (typesetting)

Why mention this? We’re talking about ‘Open’ Science!

What is ‘Open Science’?

‘Open science ... promotes transparency, accessibility, and collaboration throughout the research process. It involves making scientific knowledge, data, and methodologies openly available to the public, researchers, and other stakeholders, without restrictions or barriers.’

Source: ChatGPT 4-o

What is the ‘Scientific Record’?

‘The scientific record is crucial for the advancement of knowledge and serves as a foundation for scientific progress. It provides a means for scientists to communicate their research findings to the broader scientific community, allowing for scrutiny, validation, and further exploration.’

Source: ChatGPT 4-o.

Pragmatically ‘Open Science’:

1. Allows creation of value and advancement.
2. Allows verification of the scientific record.
3. Allows public disclosure and engagement.
4. Allows science to become accessible/transparent.

*"If I have seen further, it is by standing on the
shoulders of giants."*

Source: Isaac Newton to Robert Hooke, in February 1676

What motivates a talk about ‘Open Science’?

- Four anecdotes from my time as a doctoral student:

What motivates a talk about ‘Open Science’?

- Four anecdotes from my time as a doctoral student:
 1. Make sure your coauthors have code.

What motivates a talk about ‘Open Science’?

- Four anecdotes from my time as a doctoral student:
 1. Make sure your coauthors have code.
 2. Be prepared to dismantle bad tests of theory.

What motivates a talk about ‘Open Science’?

- Four anecdotes from my time as a doctoral student:
 1. Make sure your coauthors have code.
 2. Be prepared to dismantle bad tests of theory.
 3. Critically analyze everyone else’s data.

What motivates a talk about ‘Open Science’?

- Four anecdotes from my time as a doctoral student:
 1. Make sure your coauthors have code.
 2. Be prepared to dismantle bad tests of theory.
 3. Critically analyze everyone else’s data.
 4. Consider using open-source software.

What motivates a talk about ‘Open Science’?

- Four anecdotes from my time as a doctoral student:
 1. Make sure your coauthors have code.
 2. Be prepared to dismantle bad tests of theory.
 3. Critically analyze everyone else's data.
 4. Consider using open-source software.
- Disclaimer: presently LNL Co-ordinator for UKRN.

GARRET CHRISTENSEN | JEREMY FREESE | EDWARD MIGUEL

One of the most important parts of the book is this table:

Compared to Original	Original Procedures	Key Differences
Same Data	Verification	Reanalysis
Different Data	Direct Replication	Extension

- Maybe you've heard 'Verification' called 'Exact Replication'.
- Increasing push towards teaching concepts of 'Replication'.
- One area where Sociology has really progressed forwards.

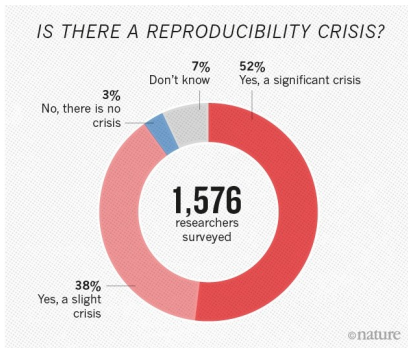
How it started:

Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • <https://doi.org/10.1371/journal.pmed.0020124>

How it's going:



‘But I do qualitative work!’

How do these ideas relate to researchers doing qualitative work?

1. Transparency in Fieldwork: Openly document fieldwork methods and decisions.
2. Open Field Notes: Share anonymized field notes for transparency.
3. Collaborative Interpretation: Invite peer or community interpretation of data.
4. Reflexive Journaling: Publish researcher reflections and biases.
5. Open Interviews: Share interview protocols and questions.

‘But I do qualitative work! (Cont.)’

6. Participant Feedback: Involve participants in data interpretation.
7. Open Coding Process: Publicize how themes and codes were developed.
8. Ethnographic Context: Provide detailed, open descriptions of research settings.
9. Transparent Ethics: Share how ethical dilemmas were navigated in the field.
10. Open Peer Review: Allow ethnographic work to be openly critiqued.

What does ‘Open Science’ look like?

1. Preprints and ‘Open Access’.
2. Publicly available data.
3. Open source code



What does ‘Open Science’ look like?

1. Preprints and ‘Open Access’.
2. Publicly available data.
3. Open source code


HARVARD
Dataverse

[Add Data](#)
[Search](#)
[About](#)
[User Guide](#)
[Support](#)
[Sign Up](#)
[Log In](#)

Deposit and share your data. Get academic credit.

Harvard Dataverse is a repository for research data. Deposit data and code here.

[Add a dataset](#)

Organize datasets and gather metrics in your own repository.

A dataverse is a container for all your datasets, files, and metadata.

[Add a dataverse](#)

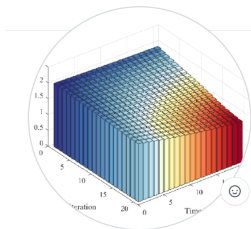
Publishing your data is easy on Harvard Dataverse!

Learn about getting started creating your own dataverse repository here.

[Getting started](#)

What does 'Open Science' look like in practice?

1. Preprints and 'Open Access'.
2. Publicly available data.
3. Open source code



Pinned

Customize your pins

[GWASReview](#)

Public ⋮

A Scientometric Review of Genome Wide Association Studies

● HTML ☆ 15 📄 7

[NHSSpend](#)

Public ⋮

A home for the NHSSpend Library

● HTML ☆ 1 📄 1

[InterModel_Vigorish](#)

Public ⋮

Introducing the InterModel Vigorish (IMV)

● Jupyter Notebook ☆ 1

[OxfordDemSci/gwasdiversitymonitor](#)

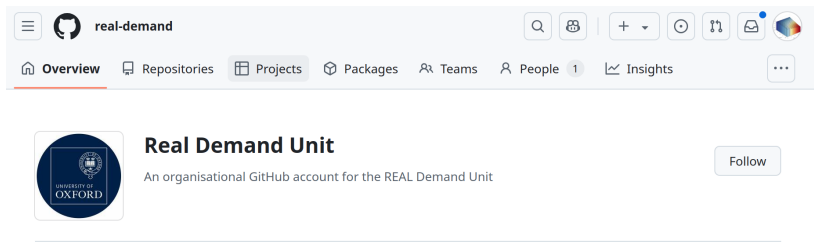
Public ⋮

Source code for the GWAS Diversity Monitor

● CSS ☆ 12 📄 1

What does ‘Open Science’ look like in practice?

1. Preprints and ‘Open Access’.
2. Publicly available data.
3. Open source code



These slides are made with an open source typesetting language!

I_make_this_slide.tex

```
\begin{frame}[]
  These slides are made with an open source typesetting language!
  \begin{center}
    \begin{scriptsize}
      \begin{minipage}{.95\linewidth}
        \begin{myenv}{guess$_\_what$__this$__does.py}
          \begin{minted}{latex}
            # Hello friends :)
          \end{minted}
        \end{myenv}
      \end{minipage}
    \end{scriptsize}
  \end{center}
\end{frame}
```

Open source statistical code can come in the form of Python (or R):

guess_what_this_does.py

```
import json, io, imageio, requests
from PIL import Image

def get_cat_bounded(url):
    api_ret = requests.get(url)
    if api_ret.status_code==200:
        text = json.loads(api_ret.text)
        while (text[0]['width']<=400 or text[0]['width']>=600) or \
            (text[0]['height']<=400 or text[0]['height']>=600):
            api_ret = requests.get(url)
            if api_ret.status_code==200:
                text = json.loads(api_ret.text)
        r = requests.get(api_text[0]['url'])
        return Image.open(io.BytesIO(r.content))

url = 'https://api.thecatapi.com/v1/images/search'
images = []
for _ in range(0, 50):
    images.append(get_cat_bounded(url))
imageio.mimsave('CATS.gif', images, fps=1)
```

Thats right: Cats!



Why show you this? Well, Python is great at lots of useful things:

1. Scientific Computing (SciPy, numpy)
2. Data visualisation (Matplotlib, seaborn)
3. Web scraping (requests)
4. Data wrangling (Pandas)

A natural choice for un/semi-structured data. Python also excels at NLP/ML, but maybe a little less with classical statistics.

Despite being qualitative researchers, I encourage learning (open source) quant tools. Let's break down the quant/qual divide!

Python is **great** at building new datasets. But why bother?

'We have been using the same data sets over and over again, and since we wanted new answers, we have been developing new econometric techniques to try to transform the data, and get more meaningful information out of them. Just having a fresh, new data set brings a whole new perspective, and I think people are starting to realize that, and gradually people are becoming more interested in data collection itself'

A. Cavallo, 15 November 2012 in Taylor et al. (2014)

Python is **great** at building new datasets. But why bother?

'We have been using the same data sets over and over again, and since we wanted new answers, we have been developing new econometric techniques to try to transform the data, and get more meaningful information out of them. Just having a fresh, new data set brings a whole new perspective, and I think people are starting to realize that, and gradually people are becoming more interested in data collection itself'

A. Cavallo, 15 November 2012 in Taylor et al. (2014)

New datasets allow fresh **insight** and are typically **open**!

5 Reasons Why Open Science Benefits You

1. Increases visibility/impact across scientific communities.
2. Promotes transparency, enhancing trust in research findings.
3. Fosters collaboration, broadening networks and opportunities.
4. Accelerates discovery by sharing data and resources openly.
5. Strengthens reproducibility, ensuring reliable/credible results.

5 Reasons Why Open Science isn't Costless

1. Time-consuming to share data, code, and protocols openly.
2. Risk of being scooped by faster competitors.
3. May expose errors; potential criticism/reputational harm.
4. Lack of incentives or recognition in some research fields.
5. Intellectual property concerns/loss of competitive advantage.

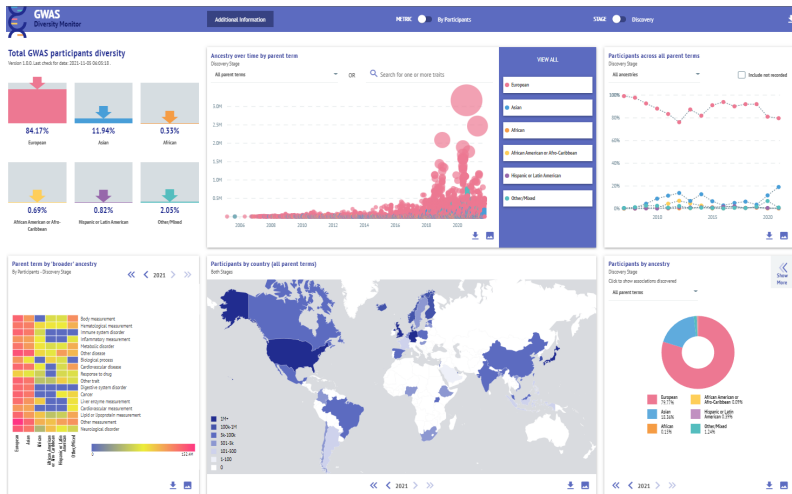
Things that Open Science is Not

1. Only for academics (it's for everyone)
2. Lowering research quality (it promotes rigorous review).
3. Only about sharing data (it encompasses all stages).
4. Removal of IP (it allows for proper attribution/licensing)
5. Requiring of all research to be open (some may remain confidential)
6. Too complex for beginners (lots of accessible tools for all levels!)

Studying Inequalities through Open Science

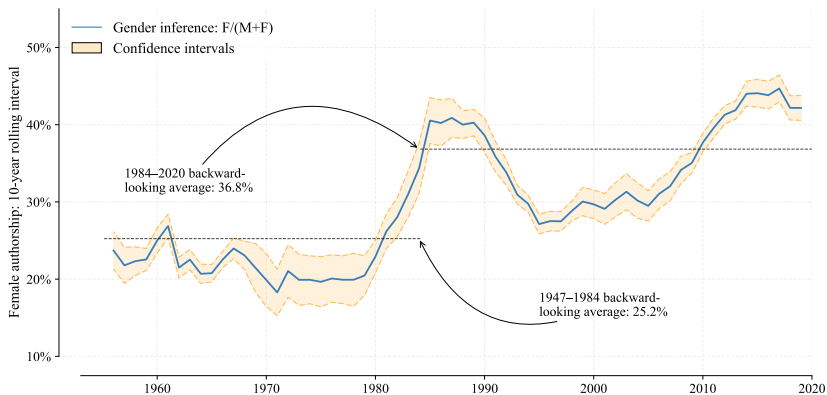
- My early scientometric work analyzed GWAS.
- It took curated lists of catalogued information from the European Bioinformatics Institute. We published two papers which:
 1. Conduct a thorough scientometric review.
 2. Launch a substantial interactive dashboard.
- Why is 'open' scientometrics good? It helps us understand what has already been done, and where inequalities remain!



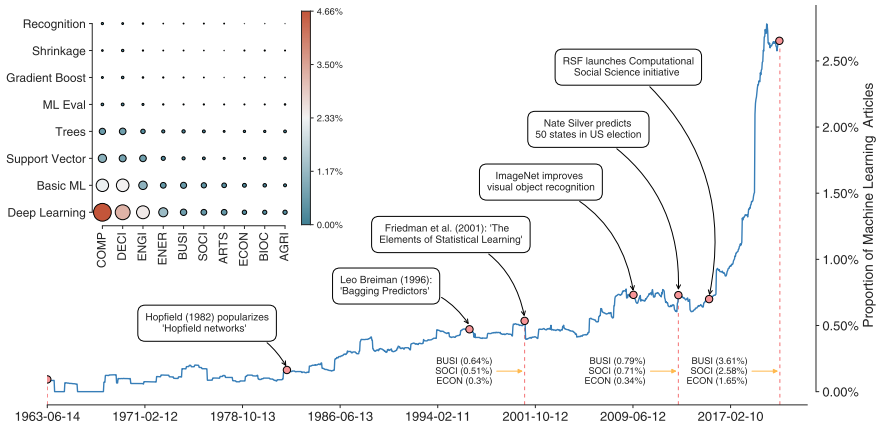


‘The GWAS Diversity Monitor Tracks Diversity in Real Time’
with Melinda C. Mills

Studying Gender with Open Science Tools



‘Population Studies at 75 Years: An empirical review’
With Melinda C. Mills



'The Rise of Machine Learning in the Academic Social Sciences'

With Dave Kirk and Mark Verhagen

Administrative Data is an invaluable resource for Good.

ADR UK (2018)

- UK leads the world in terms of ‘Openness’ of government data.
- The value of public government data is increasingly recognised.
 - Increasingly recognised by the ESRC (e.g. ADRNs).
- For better or worse:
 - I’m more interested in ‘hard to use’, semi-structured data.

Administrative Data is an invaluable resource for Good.

ADR UK (2018)

- UK leads the world in terms of ‘Openness’ of government data.
- The value of public government data is increasingly recognised.
 - Increasingly recognised by the ESRC (e.g. ADRNs).
- For better or worse:
 - I’m more interested in ‘hard to use’, semi-structured data.
 - Using curated data from the ONS SRS would be too easy.

LOBBYIST



Procurement Data

- David Cameron opened data on government contracts (2010).
- This provided a totally new type of semi-structured data.
- From from local authorities, to the NHS, to central government.
- This sounds like it might be useful for academic research...
- However, the raw data is **unbelievably** messy.

Procurement Data

- David Cameron opened data on government contracts (2010).
- This provided a totally new type of semi-structured data.
- From from local authorities, to the NHS, to central government.
- This sounds like it might be useful for academic research...
- However, the raw data is **unbelievably** messy.

If only there was a **Government Data Service** to curate it...

Procurement Data

- David Cameron opened data on government contracts (2010).
- This provided a totally new type of semi-structured data.
- From from local authorities, to the NHS, to central government.
- This sounds like it might be useful for academic research...
- However, the raw data is **unbelievably** messy.

If only there was a **Government Data Service** to curate it...

However, I am currently working with the ONS to improve collection!

Lets list a few of the problems with procurement data. Some are common to semi-structured data, some are not:

1. Some organisations don't believe in their legal obligation.
2. Every file is in a heterogeneous format in a different location.
(There are **thousands** of locations on the internet.)
3. Data commonly converted from a tabular file...

Lets list a few of the problems with procurement data. Some are common to semi-structured data, some are not:

1. Some organisations don't believe in their legal obligation.
2. Every file is in a heterogeneous format in a different location.
(There are **thousands** of locations on the internet.)
3. Data commonly converted from a tabular file... **into** a .pdf!
(This is a **direct** downgrade in the Berners-Lee 5* schema.)
4. Naturally, of course, there are no supplier UIDs.

Just one quick example, followed by what we can do...

Expenditure Over Threshold

38 / 58

Expenditure Over Threshold

38 / 58

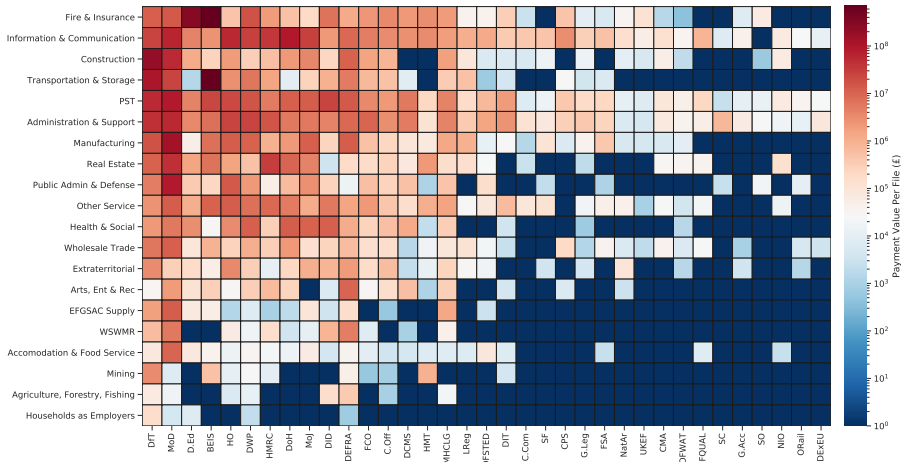
Expenditure Over Threshold

Adventures in Open and Assumption-Free Research

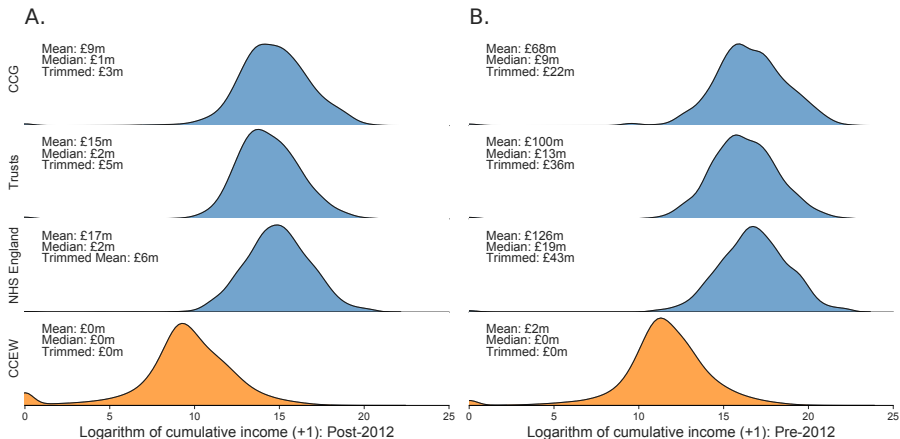
A3131

Expenditure Over Threshold

Department/Agency	City	Supply	Contract	Supplier	Transaction number	AP Amount	VAT registration number	Purchase Invoice number
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Clinic Public Sector	COMMUNITY SERVICES	115996533	666 000 000	6100071865
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Clinic Public Sector	COMMUNITY SERVICES	115996533	100 000 000	6100071861
Department of health	Norfolk	CGG	10042601	Open Gem Supplies & Ltd	COMMUNITY SERVICES	115996533	666 000 000	6100071862
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Clinic Public Sector	WYLLIE RESTAURANT	115996533	64 000 000	6100071867
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11607291	68 226 000	601008944
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING - NON ACUTE	11607291	68 226 000	601008944
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Independent Sector	COMMUNITY SERVICES	11615604	30 000 000	6100070880
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Independent Sector	ACUTE COMMISSIONING	11615604	38 526 240	6100070882
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11615604	38 526 240	6100070882
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Independent Sector	HOSPICES	11626137	419 000 000	871
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Independent Sector	ACUTE COMMISSIONING	11634001	4 440 144 000	6448267
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11634001	216 145 100	6448267
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11634001	25 672 530	48413715
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11634001	304 000 000	1000004184
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11634001	30 000 000	6100071839
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11654420	-3 663 119 340	CM00218970
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	AMBUCLANCE SERVICES	11665370	587 757 290	100111174
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11665903	768 636 000	227 480
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11665903	768 636 000	233561
Department of health	Norfolk	CGG	10042601	Cont Care Adult 100% Fully Funded	FUNDED NURSING CARE	11666982	26 912 000	2387911641
Department of health	Norfolk	CGG	10042601	Charges From CGU	RECHARGES	11666999	85 249 000	7311609994
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11671000	278 000 000	1206094
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11680115	82 568 000	600126903
Department of health	Norfolk	CGG	10042601	Hilton-Foundation Trusts	COMMUNITY SERVICES	11701702	228 968 000	278586
Department of health	Norfolk	CGG	10042601	Hilton-Foundation Trusts	COMMUNITY SERVICES	11701702	141 917 000	37932
Department of health	Norfolk	CGG	10042601	Hilton-Foundation Trusts	COMMUNITY SERVICES	11701710	807 726 000	37892
Department of health	Norfolk	CGG	10042601	Income Mx - L11	BALANCE SHEET	11702068	47 733 000	604482331
Department of health	Norfolk	CGG	10042601	Income Mx - L11-Student Loans	BALANCE SHEET	11702068	738 000	604482331
Department of health	Norfolk	CGG	10042601	National Insurance - L1-NE-PMS	BALANCE SHEET	11702068	29 114 840	604482331
Department of health	Norfolk	CGG	10042601	National Insurance - L1-NE	BALANCE SHEET	11702068	19 284 116	604482331
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11724081	213 245 000	2029126
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11724081	288 141 000	2295894
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11724081	27 100 000	2029126
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11722209	204 830 000	231384
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11724114	242 139 000	2029126
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Clinic Public Sector	CLARE RES EARTH	11725501	200 000 000	600019276
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11738262	-20 056 340	1365
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11738262	30 540 000	600019288
Department of health	Norfolk	CGG	10042601	Hilton-Foundation Trusts	COMMUNITY SERVICES	11748400	744 960 000	175451
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11748386	47 400 000	201000982
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11771720	10 000 000	263322
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11771720	2 691 952 000	1000004308
Department of health	Norfolk	CGG	10042601	Hilton-Foundation Trusts	MENTAL HEALTH CONTRACTS	11777627	25 202 000	800027387
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Independent Sector	COMMUNITY SERVICES - NON ACUTE	11777627	56 381 000	6100070880
Department of health	Norfolk	CGG	10042601	Hilton-Foundation Trusts	MENTAL HEALTH CONTRACTS	11778202	1 728 240 000	378917
Department of health	Norfolk	CGG	10042601	Hilton-Foundation Trusts	MENTAL HEALTH CONTRACTS	11801176	30 000 000	378908
Department of health	Norfolk	CGG	10042601	CLARE RES EARTH	PROPERTY SERVICES LTD	11801176	11 000 000	484138351
Department of health	Norfolk	CGG	10042601	Hilton-Foundation Trusts	CHILD AND ADOLESCENT MENTAL HEALTH	11818020	258 781 000	378905
Department of health	Norfolk	CGG	10042601	Hilton-Foundation Trusts	CHILD AND ADOLESCENT MENTAL HEALTH	11818020	58 506 000	378904
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Clinic Public Sector	HEALTHY SUPPORTIVE LIFE	11811900	118 100 000	600019283
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Independent Sector	DAY THERAPIES	11815190	276 741 000	589
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Clinic Public Sector	DAY THERAPIES	11815944	25 000 000	600019283
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11824000	1 860 931 000	6100071837
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Clinic Public Sector	LEARNING DIFFICULTIES	11854467	250 000 000	6100071247
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11857126	1 100 000 000	6100071247
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Independent Sector	COMMUNITY SERVICES	11863491	66 592 390	6100071247
Department of health	Norfolk	CGG	10042601	Healthcare Sln Res CGG	NHS 111	11923323	71 147 480	1002102149
Department of health	Norfolk	CGG	10042601	ClinicalMedica-Independent Sector	COMMUNITY SERVICES	11923323	26 703 000	6100071247
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11932379	1 137 848 900	234487
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11931340	1 455 963 000	664923417
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11933581	1 133 848 900	664923417
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11933581	-129 548 000	84686005
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11951210	1 187 569 000	232736
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11951210	200 000 000	CM00218929
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11951210	229 254 290	232340
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	PLANNED CARE	11951210	-411 000 000	CM00218929
Department of health	Norfolk	CGG	10042601	Cont Care Adult 100% Fully Funded	CHC AD FULL FUND PERKS HLTH RISK	11961590	209 532 970	6100071865
Department of health	Norfolk	CGG	10042601	Heath Sln Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	11962539	68 226 000	601008944

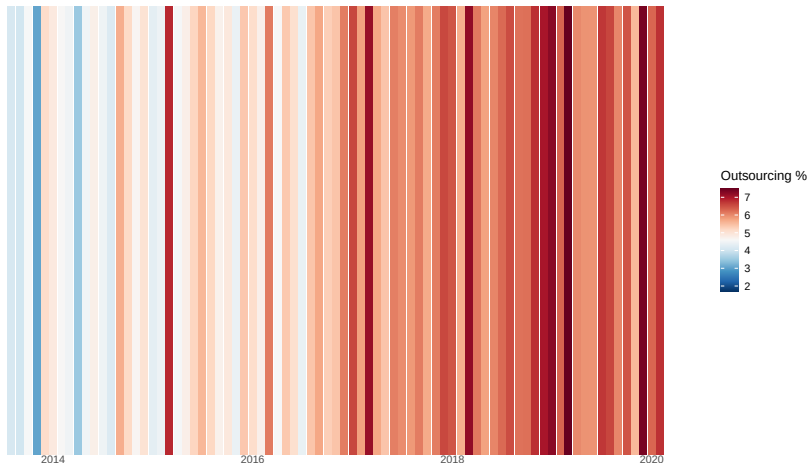


‘Tools for Transparency in Central Government Spending’



‘The Role of Non-Profits in Public Health Service Provision’ With John Mohan

Creeping Privatisation in England's NHS

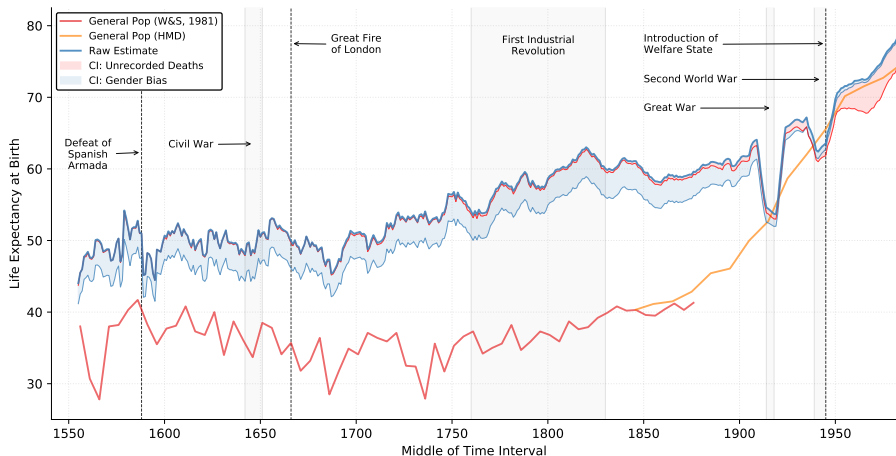


Source: Raw CCG payments approximately matched to Companies House and the Charity Commission, then manually verified

‘Creeping Commercialization in the NHS’

By Ben Goodair and Aaron Reeves



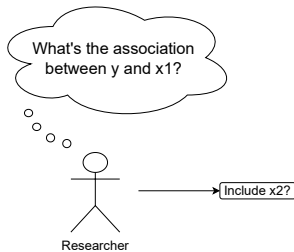


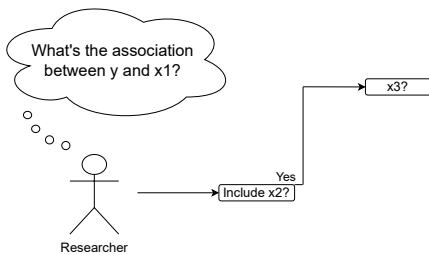
‘The Legacy of Longevity: Inequalities in UK Life Expectancy 1500-2016’ with Aaron Reeves, Felix Tropsch and Darryl Lundy

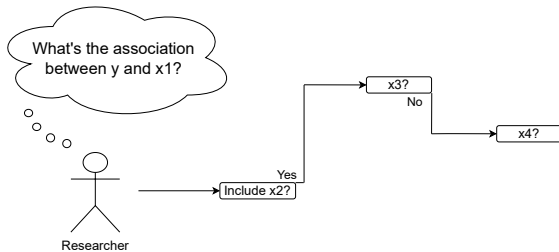
- There are close links between ‘Open Science’ and traditional modeling paradigms!
- By **opening** research, we can examine the certainty of statistical results.
- Does β_1 really equal 7.12452?

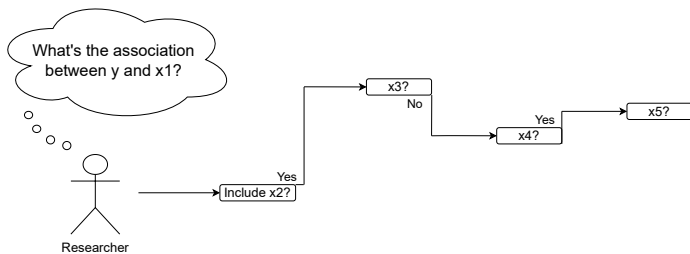
- There are close links between ‘Open Science’ and traditional modeling paradigms!
- By **opening** research, we can examine the certainty of statistical results.
- Does β_1 really equal 7.12452?
- How can we "take the ‘con’ out of econometrics"?

- There are close links between ‘Open Science’ and traditional modeling paradigms!
- By **opening** research, we can examine the certainty of statistical results.
- Does β_1 really equal 7.12452?
- How can we "take the ‘con’ out of econometrics"?
- Transparency and more computational procedures!

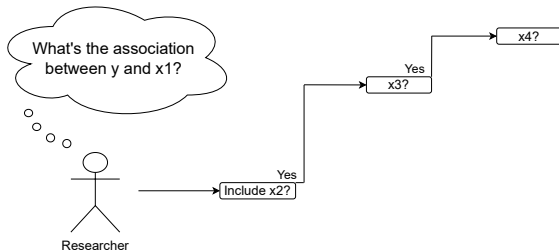


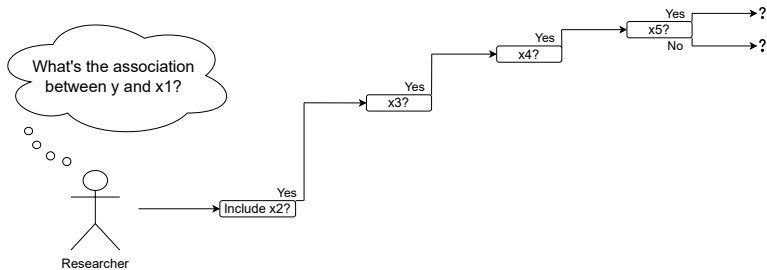


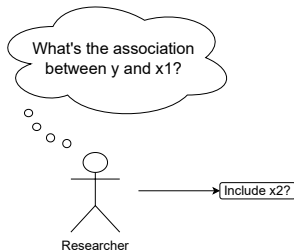


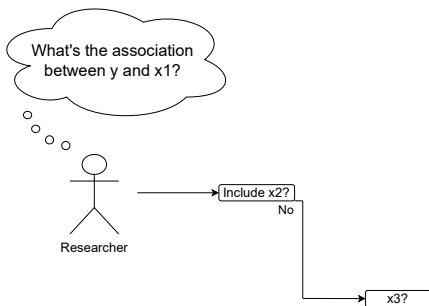


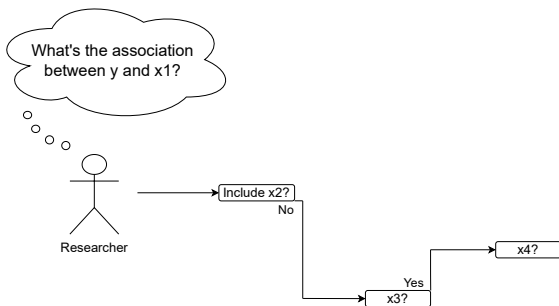


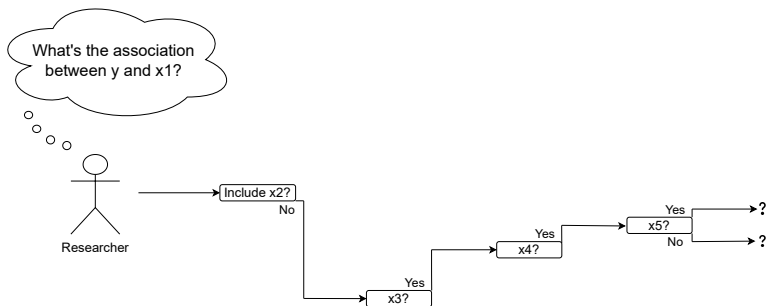


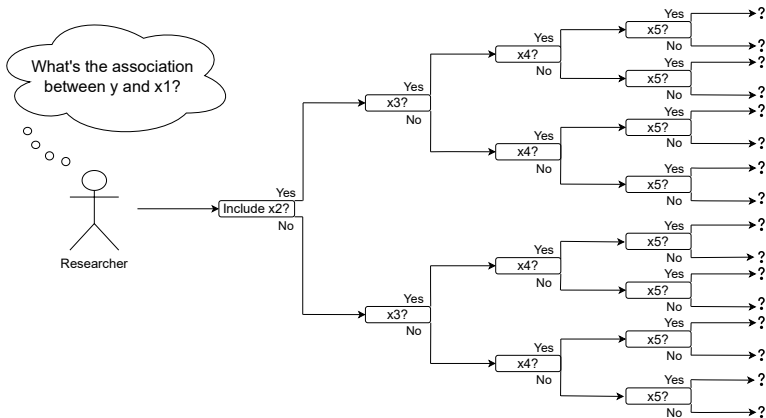




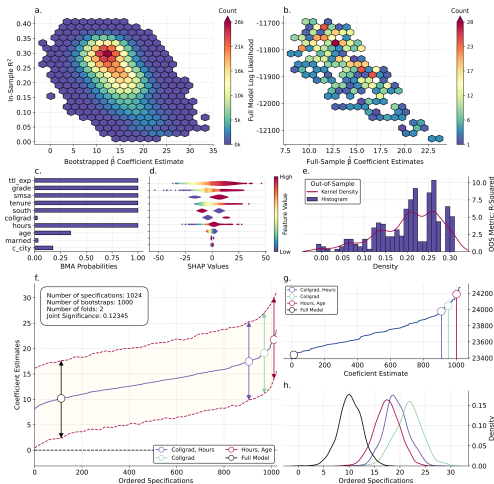








Our Attempt at Multiversal Analysis



1. Make modelling uncertainty more transparent!
2. Incorporate all multiverse tools in one place!
3. Allow researchers to do more rigorous quantitative research!



A simple, minimal website for RobustiPy!

[View the Project on GitHub here!](#)

[View the Python Documentation here!](#)

[View the RobustiPy working paper here!](#)

Welcome to the RobustiPy homepage!

RobustiPy is an efficient multiversal library with model selection, averaging, resampling and out-of-sample analysis. It analyses various output spaces, in addition to the control variable space (e.g. multiple dependent variables, estimands of interest, etc). Developed for Python and R, it is designed to be both accessible and computationally efficient.

Note: This site (and project!) are a work in progress!

Installation

Installation is as simple (in Python) as:

```
pip install robustipy
```

and – potentially in the future for R – as:

```
install.packages("robustipy")
```

Examples: A series of toy examples with which might be useful to get you up and running can be found [here](#). A series of empirical examples which might be useful can be found [here](#).

Hackathon: We are holding a hackathon regarding RobustiPy in Oxford during June 2024. Participants can click [here](#) for further information on the day's activities!

Inspiration: We found great inspiration in the following three papers, which you might also enjoy reading:

Estimating the Cost of Informal Care: A novel two-stage approach to individual synthetic control

Maria Petrillo^{1,2*}, Daniel Valdenegro Ibarra^{1,3*}, Charles Rahal^{1,3,4}, Yanan Zhang^{1,5,6}, Gwilym Pryce^{1,7}, and Matthew R. Bennett^{1,8}

¹ESRC Centre for Care

²CIRCLE, University of Sheffield

³Leverhulme Centre for Demographic Science, University of Oxford

⁴Nuffield College, University of Oxford

⁵Oxford Institute of Population Aging, University of Oxford

⁶University College, University of Oxford

⁷Department of Economics, University of Sheffield

⁸School of Social Policy, University of Birmingham

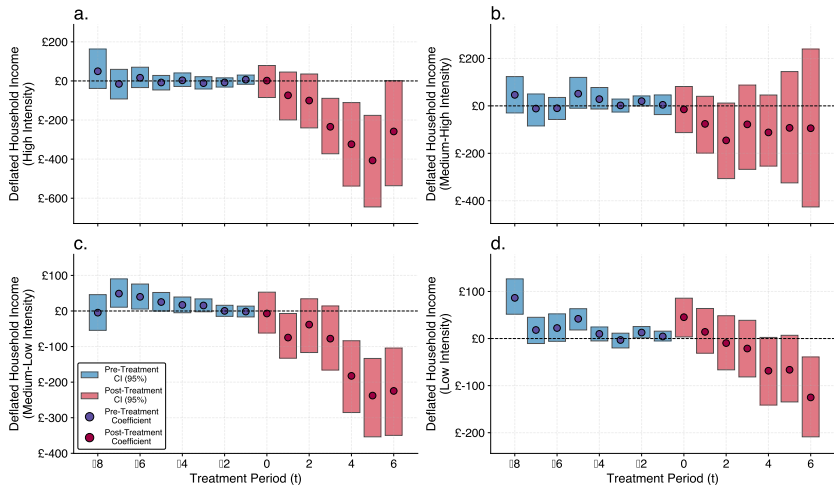
*Denotes joint lead authorship

September 19, 2024

Abstract

Informal carers provide the majority of care for people living with challenges related to older age, long-term illness, or disability. However, the care they provide carries a potentially significant income penalty for caregivers, a factor that is largely overlooked in both the economics literature and in policy discourse. Leveraging data from the UK Household Longitudinal Study, this paper provides the first robust causal estimates of the caregiving income penalty using a novel individual synthetic control method that accounts for unit-level heterogeneity in post-treatment trajectories over time. Our baseline estimates identify an average relative income gap of up to 45%, with an average decrease of £162 in monthly income, peaking at £192 per month after 4 years, based on the difference between informal carers providing the highest-intensity of care and their synthetic counterparts. We find that the income penalty is more pronounced for women than for men, and varies by ethnicity and age.

Keywords: *Causal Methods, Informal Care, Social Care, Longitudinal Analysis.*



 **costofcare** Public

 Edit Pins  Unwatch 2  Fork 0  Star 2

 main 2 Branches 0 Tags

 Add file  Code

 **MPetrillo25** Create psm e06ccf1 · 3 days ago 87 Commits

	.vscode	small analysis changes	6 months ago
	Stata_code	Create psm	3 days ago
	figs	changes made up to paa 2024	4 months ago
	output/figures	changes made up to paa 2024	4 months ago
	outputs	changes made up to paa 2024	4 months ago
	replication_attempt	Reorganising to move stuff out of root	8 months ago
	src	changes made up to paa 2024	4 months ago
	.gitignore	Reorganising to move stuff out of root	8 months ago
	README.md	Update README.md	2 years ago

 **README**

Cost of Care

This repository documents our work on calculating the cost of providing informal care. It takes an intersectional, comparative approach based on the [Individual Synthetic Control Method](#) of [Vagni and Breen \(2021\)](#).

No description, website, or topics provided.

-  Readme
-  Activity
-  Custom properties
-  2 stars
-  2 watching
-  0 forks

Report repository

Releases

No releases published
[Create a new release](#)

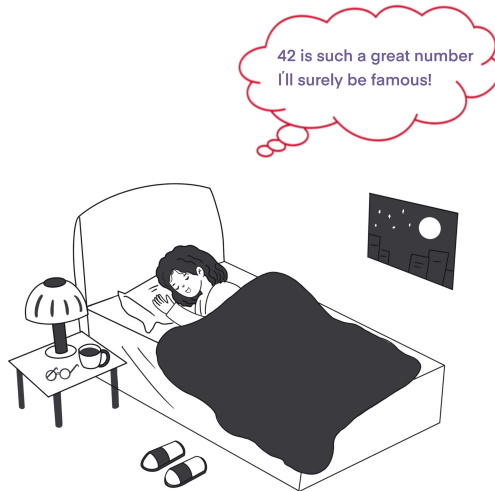
Packages

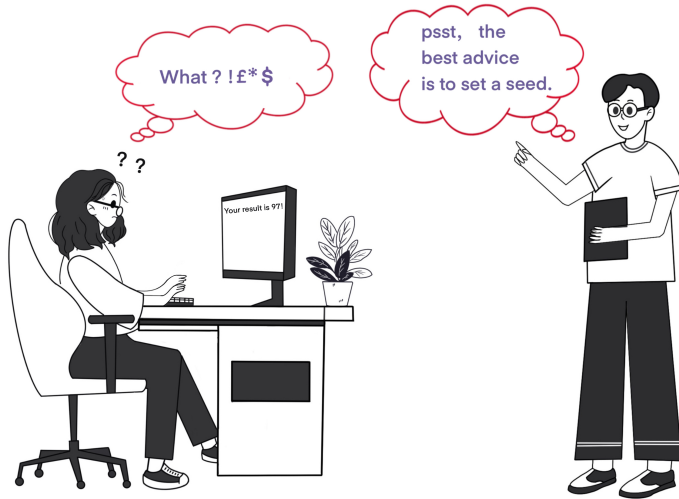
No packages published
[Publish your first package](#)

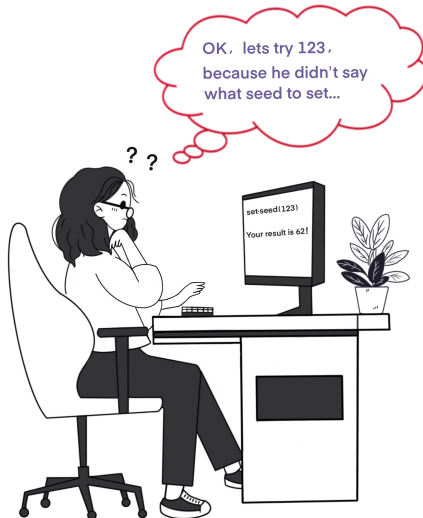
Contributors 3

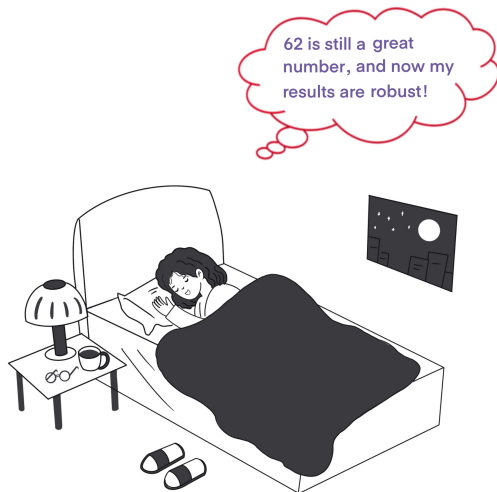
-  **dhvalden** Daniel Valdenegro Ibarra
-  **crahal**
-  **MPetrillo25**

















Pseudo-Random Number Generation (Cont.)

- We want to consider possible variation independent of arbitrary variation in seed choice.

Pseudo-Random Number Generation (Cont.)

- We want to consider possible variation independent of arbitrary variation in seed choice.
- This is an extremely important and scarcely research problem.

Pseudo-Random Number Generation (Cont.)

- We want to consider possible variation independent of arbitrary variation in seed choice.
- This is an extremely important and scarcely research problem.
- Pseudo-Random Number Generators occur **everywhere!**

Pseudo-Random Number Generation (Cont.)

- We want to consider possible variation independent of arbitrary variation in seed choice.
- This is an extremely important and scarcely research problem.
- Pseudo-Random Number Generators occur **everywhere!**
- The variation in estimand can be **huge**.

Pseudo-Random Number Generation (Cont.)

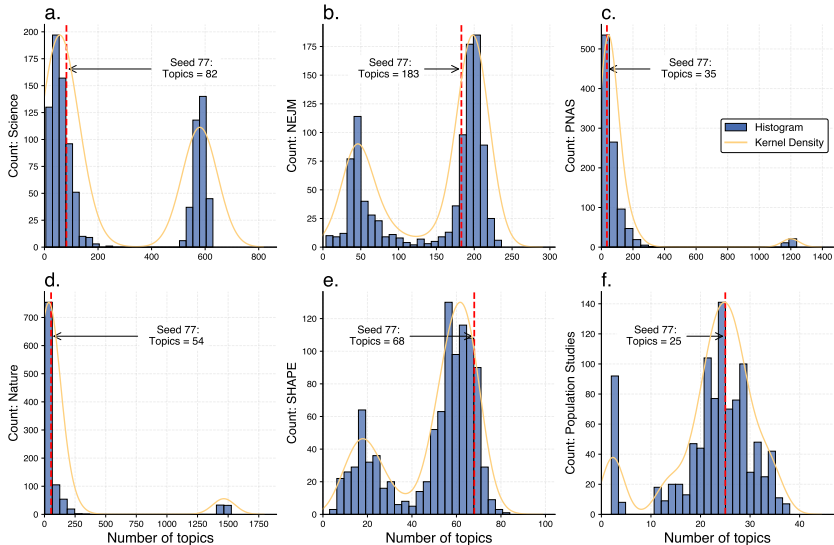
- We want to consider possible variation independent of arbitrary variation in seed choice.
- This is an extremely important and scarcely research problem.
- Pseudo-Random Number Generators occur **everywhere!**
- The variation in estimand can be **huge**.
- We bring attention to this through multiple types of replications.
 - Simulations, machine learning, topic modelling, and inferential research.

- **Problem Statement:** By setting one seed, we have no idea about the potential key variation of our estimand as a function of how random numbers were generated. This is computationally unintensive, but scientifically dour.

- **Problem Statement:** By setting one seed, we have no idea about the potential key variation of our estimand as a function of how random numbers were generated. This is computationally unintensive, but scientifically dour.
- **Solution:** Visualize the outcome space of a of large number (10k? 100k?) seeds simultaneously. This is computationally intensive, but scientifically illuminating.

- **Problem Statement:** By setting one seed, we have no idea about the potential key variation of our estimand as a function of how random numbers were generated. This is computationally unintensive, but scientifically dour.
- **Solution:** Visualize the outcome space of a of large number (10k? 100k?) seeds simultaneously. This is computationally intensive, but scientifically illuminating.
- **Replication:** We can consider whether an original result is in the tail of our distribution (or IQR?) or not.





‘From a Seed of Doubt Grows a Forest of Uncertainty’
 With Arun Frey, Jiani Yan and Mark Verhagen

Conclusion

I hope I've been able confer my excitement, enthusiasm and inspiration for:

- 'Assumption free' social science.
- 'Semi-structured' data found 'in the wild'.
- Openness and Transparency in (Care) research.
- An appreciation of how to better evaluate quantitative research, uncertainty, and the value of 'Openness'.

If just one more 'open' paper, program, design, or estimator enters scientific record, everything here will have been worth while it ☺.

Thank you for attending!

Thank you to my collaborators!

Don't hesitate to get in touch at:

charles.rahal@demography.ox.ac.uk

and

github.com/crahal!