

Adventures in Open and Assumption-Free Research

Charlie Rahal

LCDS and Nuffield College, University of Oxford

Autumn 2025



Beginning to think about Open Science

- When you entered the room, you were allocated a number!

Beginning to think about Open Science

- When you entered the room, you were allocated a number!
- Each number pertains to a question which you are going to be asked!

Beginning to think about Open Science

- When you entered the room, you were allocated a number!
- 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

- When you entered the room, you were allocated a number!
- 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'.

Thought Experiments

1. “One piece of science knowledge everyone should know?”

Thought Experiments

1. “One piece of science knowledge everyone should know?”
2. “Which science myth should be corrected?”

Thought Experiments

1. “One piece of science knowledge everyone should know?”
2. “Which science myth should be corrected?”
3. “Global problem solvable faster with open data?”

Thought Experiments

1. “One piece of science knowledge everyone should know?”
2. “Which science myth should be corrected?”
3. “Global problem solvable faster with open data?”
4. “Which discovery would change the world if shared?”

Thought Experiments

1. “One piece of science knowledge everyone should know?”
2. “Which science myth should be corrected?”
3. “Global problem solvable faster with open data?”
4. “Which discovery would change the world if shared?”

Thought Experiments (Cont.)

5. “How can scientists reach non-scientists better?”

Thought Experiments (Cont.)

5. “How can scientists reach non-scientists better?”
6. “Which field should be instantly made transparent?”

Thought Experiments (Cont.)

5. “How can scientists reach non-scientists better?”
6. “Which field should be instantly made transparent?”
7. “Everyday thing more people should know the science of?”

Thought Experiments (Cont.)

5. “How can scientists reach non-scientists better?”
6. “Which field should be instantly made transparent?”
7. “Everyday thing more people should know the science of?”
8. “How to encourage wider participation in open science?”

Thought Experiments (Cont.)

9. “Historical breakthrough faster with open science?”

Thought Experiments (Cont.)

9. “Historical breakthrough faster with open science?”
10. “If you could share one dataset with all?”

Thought Experiments (Cont.)

- 9. “Historical breakthrough faster with open science?”
- 10. “If you could share one dataset with all?”
- 11. “What risks come with open scientific methods?”

Thought Experiments (Cont.)

9. “Historical breakthrough faster with open science?”
10. “If you could share one dataset with all?”
11. “What risks come with open scientific methods?”
12. “What ethical issues arise from open scientific data?”

Learning Objectives

The remainder of this session aims to:

- Give an overview of 'Open Science' principles.
- Discuss what it is, and, importantly is not.
- Briefly mention how it applies to qualitative work.
- Despite quant focus, it is a major general (REF) debate.
- Introduce you to some of my 'Open' research work.
 - Discuss what open quant research can make possible.
- Aim: give a broader 'flavour' of the potential of Open work.

In the spirit of ‘Open Science’:

crahal.github.io/files/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)

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.

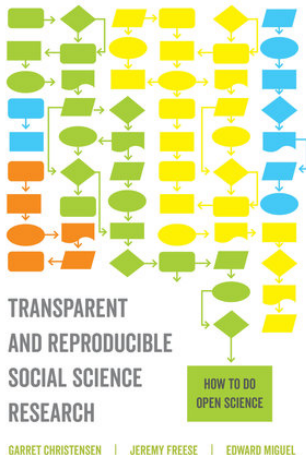
What motivates a talk about ‘Open Science’?

- 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.

A truly brilliant book:



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.

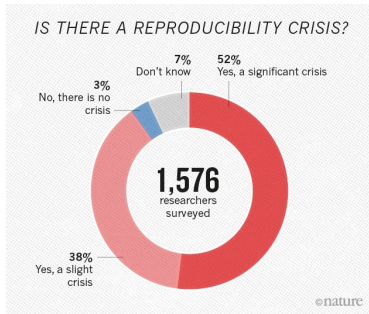
How it started:

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.

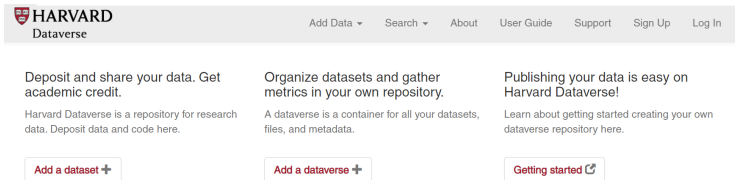
What does 'Open Science' look like?

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



The screenshot shows the Harvard Dataverse website. At the top is a navigation bar with the Harvard Dataverse logo on the left and links for 'Add Data', 'Search', 'About', 'User Guide', 'Support', 'Sign Up', and 'Log In' on the right. Below the navigation bar, the page is divided into three columns. The first column is titled 'Deposit and share your data. Get academic credit.' and contains a description of Harvard Dataverse as a repository for research data, with a button 'Add a dataset' below. The second column is titled 'Organize datasets and gather metrics in your own repository.' and contains a description of a dataverse as a container for datasets, files, and metadata, with a button 'Add a dataverse' below. The third column is titled 'Publishing your data is easy on Harvard Dataverse!' and contains a description of how to get started creating a repository, with a button 'Getting started' below.

HARVARD
Dataverse

Add Data ▾ Search ▾ About User Guide Support Sign Up Log In

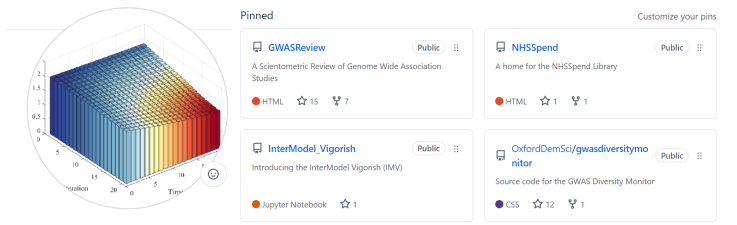
Deposit and share your data. Get academic credit.
Harvard Dataverse is a repository for research data. Deposit data and code here.
[Add a dataset](#)

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



Why Preprint?

- **Rapid dissemination** — share results without journal delay.
- **Open access** — immediate visibility for those without subscriptions.
- **Scholarly priority** — timestamped proof of contribution.
- **Community feedback** — receive critique before formal review.
- **Transparency** — aligns with open science and reproducibility ideals.

Why Preprint?

- **Rapid dissemination** — share results without journal delay.
- **Open access** — immediate visibility for those without subscriptions.
- **Scholarly priority** — timestamped proof of contribution.
- **Community feedback** — receive critique before formal review.
- **Transparency** — aligns with open science and reproducibility ideals.

Preprints accelerate communication and foster collective improvement.

Why Not Preprint?

- **Quality control** — unreviewed results may contain errors.
- **Journal policy conflicts** — some venues restrict prior circulation.
- **Public misunderstanding** — preliminary results may be misinterpreted.[0.25em]
- **Scooping concerns** — ideas exposed before peer-reviewed validation.[0.25em]
- **Citation instability** — versioning can fragment scholarly attribution.

Why Not Preprint?

- **Quality control** — unreviewed results may contain errors.
- **Journal policy conflicts** — some venues restrict prior circulation.
- **Public misunderstanding** — preliminary results may be misinterpreted.[0.25em]
- **Scooping concerns** — ideas exposed before peer-reviewed validation.[0.25em]
- **Citation instability** — versioning can fragment scholarly attribution.

Preprints trade control for speed — a calculated academic risk.

When to Preprint

- When transparency and rapid feedback outweigh reputational risk.
- When the work benefits from early collaboration or data sharing.
- When the target journal explicitly permits or encourages preprints.
- When open dissemination aligns with funder or institutional policy.

When to Preprint

- When transparency and rapid feedback outweigh reputational risk.
- When the work benefits from early collaboration or data sharing.
- When the target journal explicitly permits or encourages preprints.
- When open dissemination aligns with funder or institutional policy.

Use preprints strategically — to invite critique, not to bypass review.

Preprints and the Future of Peer Review

- Shifts validation from closed refereeing to open community evaluation.
- Encourages post-publication commentary and iterative refinement.
- Increases accountability — work is visible before formal acceptance.
- May reshape incentives toward transparency over prestige.

Preprints and the Future of Peer Review

- Shifts validation from closed refereeing to open community evaluation.
- Encourages post-publication commentary and iterative refinement.
- Increases accountability — work is visible before formal acceptance.
- May reshape incentives toward transparency over prestige.

Preprints do not replace peer review — they redefine its temporal order.

Why Use GitHub or GitLab?

- **Version control** — track, revert, and compare all changes.
- **Collaboration** — concurrent contributions without data loss.
- **Transparency** — visible research process promotes reproducibility.
- **Persistence** — repositories are archived, citable, and durable.
- **Integration** — links code, data, and documentation in one ecosystem.

Why Use GitHub or GitLab?

- **Version control** — track, revert, and compare all changes.
- **Collaboration** — concurrent contributions without data loss.
- **Transparency** — visible research process promotes reproducibility.
- **Persistence** — repositories are archived, citable, and durable.
- **Integration** — links code, data, and documentation in one ecosystem.

A repository is both a laboratory notebook and a publication record.

Scientific Reproducibility

- **Complete provenance** — commit history records every computational state.
- **Code–data linkage** — scripts, outputs, and figures stay synchronized.
- **Automated builds** — continuous integration ensures code validity.
- **Reusable workflows** — others can replicate your environment precisely.

Scientific Reproducibility

- **Complete provenance** — commit history records every computational state.
- **Code–data linkage** — scripts, outputs, and figures stay synchronized.
- **Automated builds** — continuous integration ensures code validity.
- **Reusable workflows** — others can replicate your environment precisely.

Reproducibility is not an ideal — it is a version-controlled practice.

Open Collaboration and Credit

- **Issues & pull requests** — formalise discussion and peer contribution.
- **Authorship trails** — contributions are time-stamped and attributable.
- **Community review** — invites critique before journal publication.
- **Open metrics** — forks, stars, and citations reflect visible impact.

Open Collaboration and Credit

- **Issues & pull requests** — formalise discussion and peer contribution.
- **Authorship trails** — contributions are time-stamped and attributable.
- **Community review** — invites critique before journal publication.
- **Open metrics** — forks, stars, and citations reflect visible impact.

Open repositories make scholarly contribution quantifiable and verifiable.

Long-Term Value for Science

- **Interoperability** — connect notebooks, datasets, and APIs.
- **Archival stability** — DOI assignment via Zenodo or GitLab export.
- **Pedagogical reuse** — codebases become teaching resources.
- **Open governance** — fosters collective maintenance beyond individuals.

Long-Term Value for Science

- **Interoperability** — connect notebooks, datasets, and APIs.
- **Archival stability** — DOI assignment via Zenodo or GitLab export.
- **Pedagogical reuse** — codebases become teaching resources.
- **Open governance** — fosters collective maintenance beyond individuals.

Git-based infrastructures institutionalise openness, ensuring science remains cumulative.

The Importance of a Good Text Editor

- **Precision:** Efficient syntax highlighting and linting reduce coding errors.
- **Speed:** Lightweight, responsive interfaces minimise cognitive friction.
- **Focus:** Distraction-free modes enhance deep work and code comprehension.
- **Extensibility:** Plugin ecosystems adapt the editor to your workflow.
- **Collaboration:** Modern editors (e.g. Zed) enable low-latency pair programming.

The Importance of a Good Text Editor

- **Precision:** Efficient syntax highlighting and linting reduce coding errors.
- **Speed:** Lightweight, responsive interfaces minimise cognitive friction.
- **Focus:** Distraction-free modes enhance deep work and code comprehension.
- **Extensibility:** Plugin ecosystems adapt the editor to your workflow.
- **Collaboration:** Modern editors (e.g. Zed) enable low-latency pair programming.

A good editor is not a tool — it is an environment for thought.

The Importance of a Project Manager

- I've search for years for a good project management app.
- I used to just track projects using pen and paper (and whiteboards).
- Then I used csv files.
- Then I used GoogleKeep.
- Then I used ToDoist.
- Now I use Notion and it is **glorious**.

Open Access

Why Use Linux? — I

- **Open Source:** Freely available and community-maintained — aligns with open science ideals.
- **Stability:** Long-term support (LTS) releases ensure consistent research environments.
- **Security:** Regular updates and strong package verification minimize vulnerabilities.
- **Efficiency:** Lightweight, modular design suited for data-intensive workflows.

Why Use Linux? — I

- **Open Source:** Freely available and community-maintained — aligns with open science ideals.
- **Stability:** Long-term support (LTS) releases ensure consistent research environments.
- **Security:** Regular updates and strong package verification minimize vulnerabilities.
- **Efficiency:** Lightweight, modular design suited for data-intensive workflows.

Linux embodies the reproducibility ethos: transparent, robust, and dependable.

Why Use Linux? — II

- **Compatibility:** Native support for Python, R, Git, and $\text{\LaTeX}$ out of the box.
- **Customisation:** Full control over system behaviour, packages, and kernels.
- **Performance:** Excellent resource management for high-performance and cloud computing.
- **Community:** Global ecosystem of developers, researchers, and maintainers.

Why Use Linux? — II

- **Compatibility:** Native support for Python, R, Git, and $\text{\LaTeX}$ out of the box.
- **Customisation:** Full control over system behaviour, packages, and kernels.
- **Performance:** Excellent resource management for high-performance and cloud computing.
- **Community:** Global ecosystem of developers, researchers, and maintainers.

Linux is not just an operating system — it is an open research platform.

Why Use $\text{\LaTeX}$? — I

- **Precision:** Mathematical notation typeset with professional clarity.
- **Structure:** Separates content from presentation, enforcing consistency.
- **Automation:** Cross-references, bibliographies, and numbering handled automatically.
- **Stability:** Plain-text source ensures long-term reproducibility and version control.

Why Use $\text{\LaTeX}$? — I

- **Precision:** Mathematical notation typeset with professional clarity.
- **Structure:** Separates content from presentation, enforcing consistency.
- **Automation:** Cross-references, bibliographies, and numbering handled automatically.
- **Stability:** Plain-text source ensures long-term reproducibility and version control.

LaTeX transforms writing into a computational act of composition.

Why Use L^AT_EX? — II

- **Interoperability:** Works seamlessly with Git, Python, and R workflows.
- **Portability:** Compiles identically across systems and decades.
- **Professionalism:** Journal-ready typography with minimal manual styling.
- **Extensibility:** Packages for slides, algorithms, and data visualisation.
- **Community:** Open source, continually refined by researchers and educators.

Why Use L^AT_EX? — II

- **Interoperability:** Works seamlessly with Git, Python, and R workflows.
- **Portability:** Compiles identically across systems and decades.
- **Professionalism:** Journal-ready typography with minimal manual styling.
- **Extensibility:** Packages for slides, algorithms, and data visualisation.
- **Community:** Open source, continually refined by researchers and educators.

LaTeX is not merely a format — it is the language of academic precision.

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 code can come in 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.

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!

Python is General-Purpose yet Expressive.

- The same language powers *scientific computing*, *web development*, *machine learning*, and *scripting*.
- Syntax is concise but expressive — high-level abstractions for mathematics, I/O, and logic.
- This breadth makes Python ideal for cross-domain research, from econometrics to computational linguistics.

Python is General-Purpose yet Expressive.

- The same language powers *scientific computing*, *web development*, *machine learning*, and *scripting*.
- Syntax is concise but expressive — high-level abstractions for mathematics, I/O, and logic.
- This breadth makes Python ideal for cross-domain research, from econometrics to computational linguistics.

“Readability counts.” — The Zen of Python

Python's Scientific Ecosystem is Mature

example_numerical_simulation.py

```
import numpy as np

def monte_carlo_pi(n=10_000_000):
    pts = np.random.rand(n, 2)
    inside = np.sum(np.sum(pts**2, axis=1) < 1)
    return 4 * inside / n

print(monte_carlo_pi())
```


Python's Scientific Ecosystem is Mature

example_numerical_simulation.py

```
import numpy as np

def monte_carlo_pi(n=10_000_000):
    pts = np.random.rand(n, 2)
    inside = np.sum(np.sum(pts**2, axis=1) < 1)
    return 4 * inside / n

print(monte_carlo_pi())
```

High-performance numerical computation via NumPy and SciPy provides reproducible, vectorised results — a cornerstone of modern empirical analysis.

Reproducibility through Open Source

- Python encourages transparent workflows: notebooks, plain-text scripts, and open repositories.
- Dependencies are declared and environment-managed (pip, conda).
- Version control integration (e.g. GitHub) allows peer review and collaborative improvement.

Reproducibility through Open Source

- Python encourages transparent workflows: notebooks, plain-text scripts, and open repositories.
- Dependencies are declared and environment-managed (pip, conda).
- Version control integration (e.g. GitHub) allows peer review and collaborative improvement.

Open tools produce open science: others can verify, extend, and build upon your results.

Python Interfaces with Other Languages.

hybrid_computation.py

```
import numpy as np
from numba import njit

@njit
def fast_square_sum(a):
    s = 0.0
    for x in a:
        s += x * x
    return s

data = np.random.rand(1_000_000)
print(fast_square_sum(data))
```

Python Interfaces with Other Languages.

hybrid_computation.py

```
import numpy as np
from numba import njit

@njit
def fast_square_sum(a):
    s = 0.0
    for x in a:
        s += x * x
    return s

data = np.random.rand(1_000_000)
print(fast_square_sum(data))
```

Through bindings such as Cython, Numba, or PyTorch C++, one gains both expressivity and speed — mathematical clarity without sacrificing performance.

A Language for Thought.

- Python's design promotes logical reasoning: indentation enforces structure, and its functional features (e.g. comprehensions, generators) mirror mathematical expression.
- The community's norms (PEP 8, docstrings, testing) elevate code to readable, shareable research artifacts.
- Its ubiquity in academia and industry forms a shared technical language across disciplines.

A Language for Thought.

- Python's design promotes logical reasoning: indentation enforces structure, and its functional features (e.g. comprehensions, generators) mirror mathematical expression.
- The community's norms (PEP 8, docstrings, testing) elevate code to readable, shareable research artifacts.
- Its ubiquity in academia and industry forms a shared technical language across disciplines.

Mathematical clarity, computational power, and openness — all in one language.

R is Purpose-Built for Statistics.

- Developed by and for statisticians — statistical modelling, inference, and visualization are native features.
- The syntax aligns with mathematical notation: formula interfaces ($y \sim x$) generalise easily to linear, generalized, and mixed models.
- Direct integration with probability distributions and hypothesis testing makes R a natural tool for empirical work.

R is Purpose-Built for Statistics.

- Developed by and for statisticians — statistical modelling, inference, and visualization are native features.
- The syntax aligns with mathematical notation: formula interfaces ($y \sim x$) generalise easily to linear, generalized, and mixed models.
- Direct integration with probability distributions and hypothesis testing makes R a natural tool for empirical work.

“R is not just a language, but a statistical environment.”

Comprehensive Modelling Capabilities.

linear_model_example.R

```
data(mtcars)
fit <- lm(mpg ~ wt + hp, data = mtcars)
summary(fit)
```

Comprehensive Modelling Capabilities.

linear_model_example.R

```
data(mtcars)
fit <- lm(mpg ~ wt + hp, data = mtcars)
summary(fit)
```

Linear, generalized linear, mixed-effects, time-series, and survival models are all natively supported, often requiring only a single formula change — an exceptionally coherent statistical syntax.

Elegant and Declarative Visualisation.

- The `ggplot2` package implements Wilkinson's Grammar of Graphics, providing a principled framework for constructing complex plots.
- Layers, aesthetics, and scales are composed declaratively rather than procedurally.
- Visualization and analysis coexist naturally, enabling exploratory data analysis as a statistical act.

Elegant and Declarative Visualisation.

- The `ggplot2` package implements Wilkinson's Grammar of Graphics, providing a principled framework for constructing complex plots.
- Layers, aesthetics, and scales are composed declaratively rather than procedurally.
- Visualization and analysis coexist naturally, enabling exploratory data analysis as a statistical act.

Graphics in R are not merely decorative; they are inferential.

Literate Programming.

reproducible_report.Rmd

```
library(ggplot2)  
qplot(wt, mpg, data = mtcars)
```

Literate Programming.

reproducible_report.Rmd

```
library(ggplot2)  
qplot(wt, mpg, data = mtcars)
```

R Markdown and `knitr` integrate code, text, and output — a single document can perform computation, display results, and describe interpretation. *A complete research pipeline in one file.*

Community, Packages, and Academic Integrity.

- CRAN hosts > 20,000 peer-reviewed packages, covering econometrics, genomics, Bayesian inference, and causal analysis.
- The academic orientation ensures statistical rigour and transparent documentation.
- RStudio (Posit) and the tidyverse have unified standards for reproducibility and pedagogy.

Community, Packages, and Academic Integrity.

- CRAN hosts > 20,000 peer-reviewed packages, covering econometrics, genomics, Bayesian inference, and causal analysis.
- The academic orientation ensures statistical rigour and transparent documentation.
- RStudio (Posit) and the tidyverse have unified standards for reproducibility and pedagogy.

R is the lingua franca of data analysis — mathematically precise, statistically principled, and open.

R vs Python

R

- Built for statistics
- Formula syntax:
 $y \sim x + z$
- Rich academic packages
- ggplot2 for elegant graphics
- Ideal for inference & reporting

Python

- General-purpose language
- Clean, expressive syntax
- Vast ML/AI ecosystem
- Strong in automation & integration
- Ideal for scalable analysis

R vs Python

R

- Built for statistics
- Formula syntax:
 $y \sim x + z$
- Rich academic packages
- ggplot2 for elegant graphics
- Ideal for inference & reporting

Python

- General-purpose language
- Clean, expressive syntax
- Vast ML/AI ecosystem
- Strong in automation & integration
- Ideal for scalable analysis

Together: theory meets production, inference meets scale.

Open-Source Tools for Qualitative Research I

- **Taguette** — lightweight, browser-based coding of text (PDFs, Word, etc.).
- **QualCoder** — full-featured CAQDAS: coding, memos, audio/video support.
- **RQDA** (R-based) — integrates qualitative coding with reproducible R workflows.
- **CATMA** — collaborative textual annotation for discourse and literary analysis.
- **Weft QDA** — minimalist but powerful open tool for structured text coding.

Open-Source Tools for Qualitative Research I

- **Taguette** — lightweight, browser-based coding of text (PDFs, Word, etc.).
- **QualCoder** — full-featured CAQDAS: coding, memos, audio/video support.
- **RQDA** (R-based) — integrates qualitative coding with reproducible R workflows.
- **CATMA** — collaborative textual annotation for discourse and literary analysis.
- **Weft QDA** — minimalist but powerful open tool for structured text coding.

Open CAQDAS tools bring transparency and longevity to qualitative projects.

Open-Source Tools for Qualitative Research II

- **Audacity** — open audio editor for cleaning interviews and focus groups.
- **ELAN** — detailed annotation of audio/video for conversation and gesture analysis.
- **oTranscribe / Parlatype** — assist manual transcription with keyboard shortcuts.
- **OpenRefine** — cleaning messy CSV exports of coded segments or survey data.
- **Praat** — acoustic analysis and annotation of speech recordings.

Open-Source Tools for Qualitative Research II

- **Audacity** — open audio editor for cleaning interviews and focus groups.
- **ELAN** — detailed annotation of audio/video for conversation and gesture analysis.
- **oTranscribe / Parlatype** — assist manual transcription with keyboard shortcuts.
- **OpenRefine** — cleaning messy CSV exports of coded segments or survey data.
- **Praat** — acoustic analysis and annotation of speech recordings.

From raw sound to structured text, open tools keep your pipeline inspectable.

Open-Source Tools for Qualitative Research III

- **Zotero** — open reference manager for organising literature and PDFs.
- **Obsidian / Joplin** — markdown note-taking for reflexive journals and memos.
- **Git + GitHub/GitLab** — version control for codebooks, transcripts, and analyses.
- **L^AT_EX** or **R Markdown** — reproducible reports combining text, tables, and figures.
- **Quarto** — unified open framework linking text, computation, and documentation.

Open-Source Tools for Qualitative Research III

- **Zotero** — open reference manager for organising literature and PDFs.
- **Obsidian / Joplin** — markdown note-taking for reflexive journals and memos.
- **Git + GitHub/GitLab** — version control for codebooks, transcripts, and analyses.
- **L^AT_EX** or **R Markdown** — reproducible reports combining text, tables, and figures.
- **Quarto** — unified open framework linking text, computation, and documentation.

Open ecosystems make your qualitative work portable, auditable, and shareable.

What's the issue with using proprietary, licensed software?

What's the issue with proprietary software?

- **Opacity:** Source code is closed — methods cannot be verified.
- **Reproducibility:** Results depend on inaccessible versions.
- **Access:** Licensing costs restrict equality.
- **Longevity:** Projects risk obsolescence when vendors discontinue support.
- **Lock-in:** Data formats and dependencies reduce researcher autonomy.

What's the issue with proprietary software?

- **Opacity:** Source code is closed — methods cannot be verified.
- **Reproducibility:** Results depend on inaccessible versions.
- **Access:** Licensing costs restrict equality.
- **Longevity:** Projects risk obsolescence when vendors discontinue support.
- **Lock-in:** Data formats and dependencies reduce researcher autonomy.

Closed tools close science — openness sustains verification and progress.

Open-Source Alternatives Solve Issues

- **Transparency:** Source code and methods are fully inspectable and auditable.
- **Reproducibility:** Identical results can be regenerated anywhere, anytime.
- **Equity:** Free access enables global participation in scientific work.
- **Sustainability:** Community maintenance ensures long-term usability.
- **Interoperability:** Open standards prevent vendor lock-in and promote integration.

Open-Source Alternatives Solve Issues

- **Transparency:** Source code and methods are fully inspectable and auditable.
- **Reproducibility:** Identical results can be regenerated anywhere, anytime.
- **Equity:** Free access enables global participation in scientific work.
- **Sustainability:** Community maintenance ensures long-term usability.
- **Interoperability:** Open standards prevent vendor lock-in and promote integration.

Open software keeps science verifiable, equitable, and enduring.

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!)

Generative AI Facilitates Open Science

- Summarizes open-access research into multilingual digests.

Generative AI Facilitates Open Science

- Summarizes open-access research into multilingual digests.
- Generates synthetic datasets; protect privacy/enables sharing.

Generative AI Facilitates Open Science

- Summarizes open-access research into multilingual digests.
- Generates synthetic datasets; protect privacy/enables sharing.
- Automates open-source workflows, enhancing reproducibility.

Generative AI Facilitates Open Science

- Summarizes open-access research into multilingual digests.
- Generates synthetic datasets; protect privacy/enables sharing.
- Automates open-source workflows, enhancing reproducibility.
- Builds dynamic repositories standardizing contributions.

Generative AI Facilitates Open Science

- Summarizes open-access research into multilingual digests.
- Generates synthetic datasets; protect privacy/enables sharing.
- Automates open-source workflows, enhancing reproducibility.
- Builds dynamic repositories standardizing contributions.
- Translates complex findings into accessible language.

GDPR and Open Science

1. Protects individuals' personal research data

GDPR and Open Science

1. Protects individuals' personal research data
2. Limits sharing of identifiable human information

GDPR and Open Science

1. Protects individuals' personal research data
2. Limits sharing of identifiable human information
3. Requires informed consent for data use

GDPR and Open Science

1. Protects individuals' personal research data
2. Limits sharing of identifiable human information
3. Requires informed consent for data use
4. Mandates anonymisation before open release

GDPR and Open Science

1. Protects individuals' personal research data
2. Limits sharing of identifiable human information
3. Requires informed consent for data use
4. Mandates anonymisation before open release
5. Imposes strict rules on sensitive categories

GDPR and Open Science (Cont.)

1. Determines which datasets can be shared openly

GDPR and Open Science (Cont.)

1. Determines which datasets can be shared openly
2. Promotes secure and restricted data access

GDPR and Open Science (Cont.)

1. Determines which datasets can be shared openly
2. Promotes secure and restricted data access
3. Balances scientific openness with privacy rights

GDPR and Open Science (Cont.)

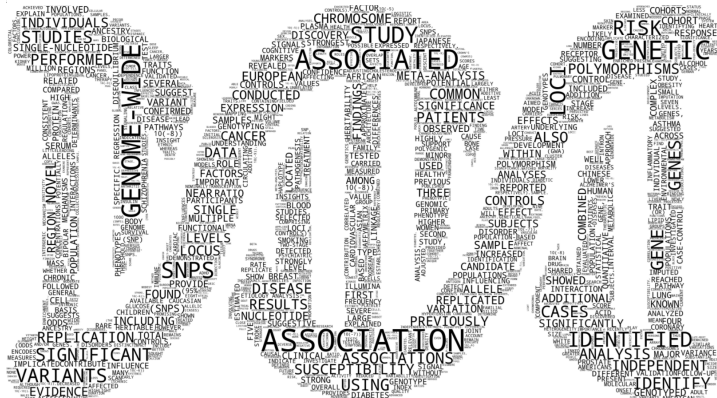
1. Determines which datasets can be shared openly
2. Promotes secure and restricted data access
3. Balances scientific openness with privacy rights
4. Varies across jurisdictions, complicating collaboration

GDPR and Open Science (Cont.)

1. Determines which datasets can be shared openly
2. Promotes secure and restricted data access
3. Balances scientific openness with privacy rights
4. Varies across jurisdictions, complicating collaboration
5. Requires researchers to plan for compliance early

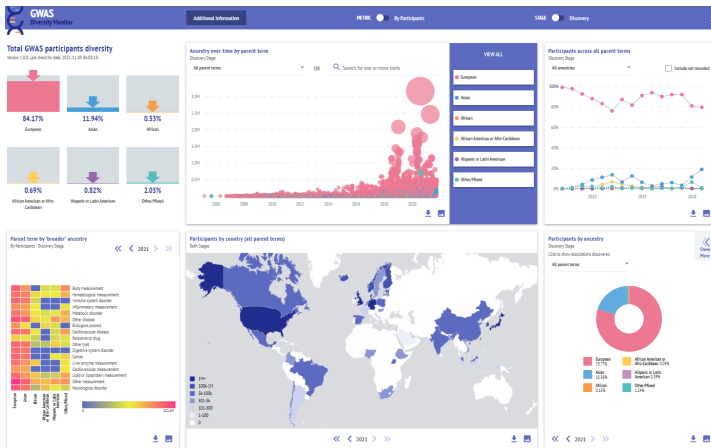
Studying Inequalities through Open Science

- My early scientometric work analyzed GWAS.
- It took curated lists of catalogued information from the European Bioinformatics Institute. We published two papers:
 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 opportunities remain!



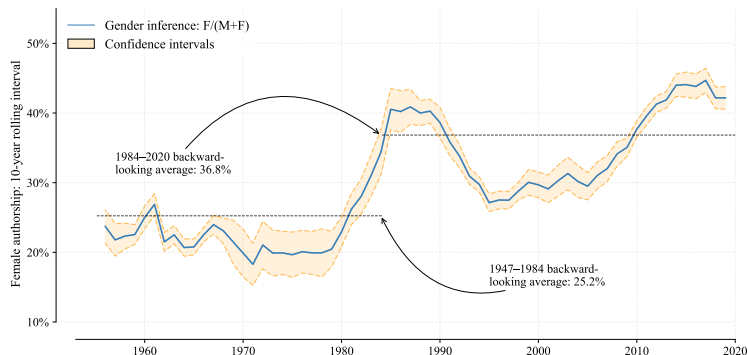
'The Scientometrics of Genome Wide Association Studies'

With Melinda C. Mills

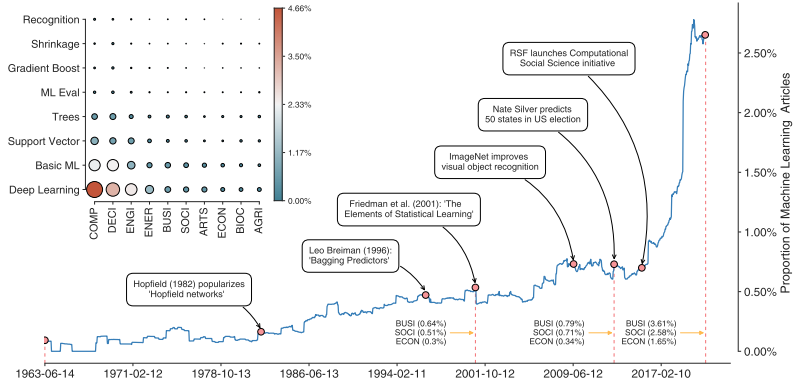


‘The GWAS Diversity Monitor Tracks Diversity in Real Time’ with Melinda C. Mills

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.

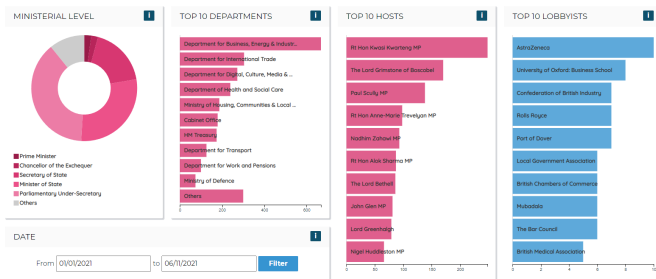
ADR UK (2018)

- 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.



openaccess.transparency.org.uk

With Ian Knowles, Paulo Serôdio, David Stuckler and TI UK

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

[illegible]

Expenditure Over Threshold

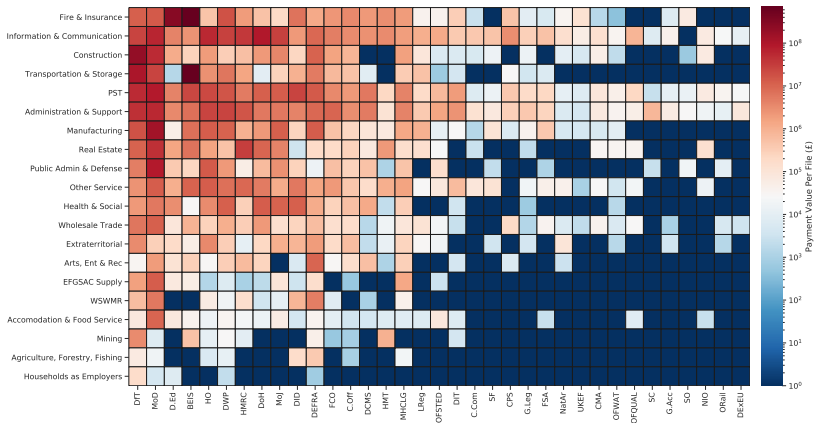
[illegible]

Expenditure Over Threshold

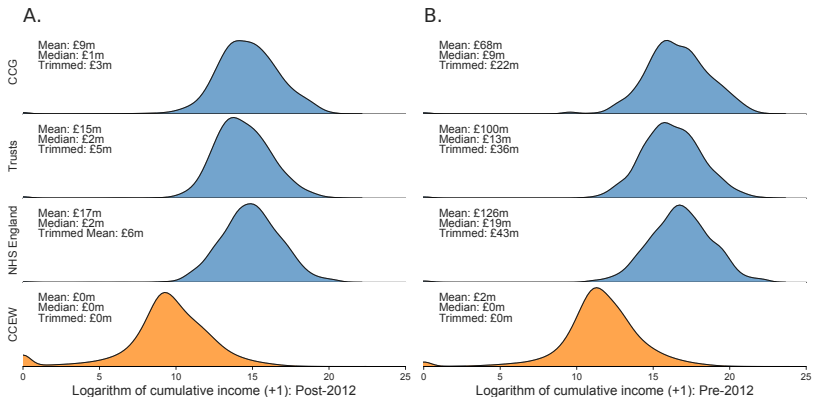
[illegible]

Expenditure Over Threshold

[illegible]



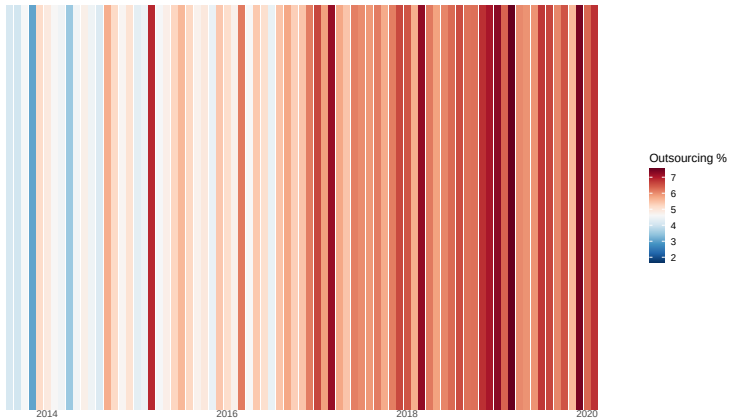
‘Tools for Transparency in Central Government Spending’



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

With John Mohan

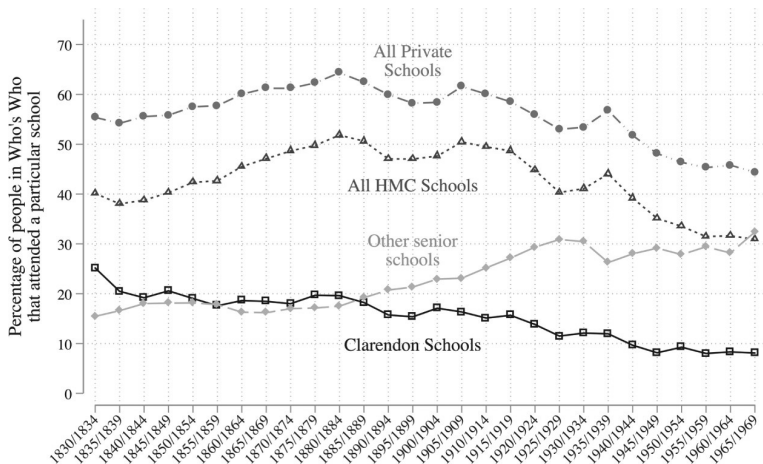
Creeping Privatisation in England's NHS



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

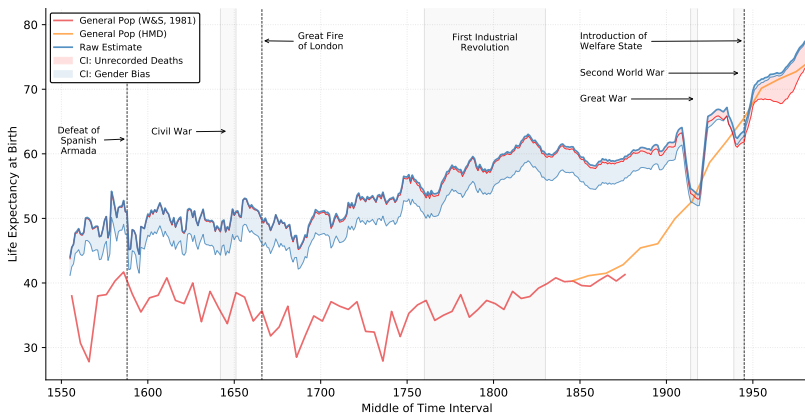
‘Creeping Commercialization in the NHS’

By Ben Goodair and Aaron Reeves



‘The decline and persistence of the old boy’

With Aaron Reeves, Sam Friedman. and Magne Flemmen



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

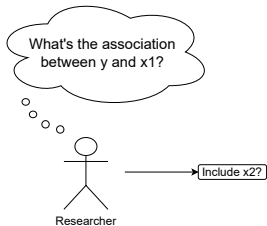
The idea is to:

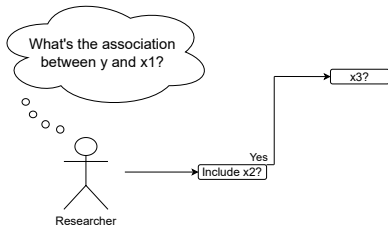
- 78 / 93

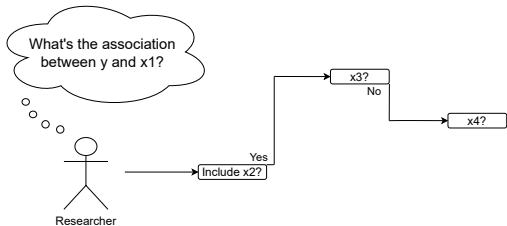
- There are close links between ‘Open Science’ and traditional modeling paradigms!
- By **opening** research, we can examine the certainty of statistical results.
- Does β_1 really equal 7.12452?

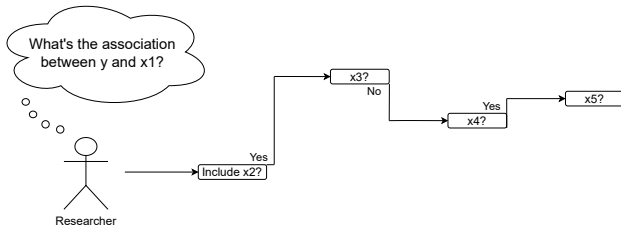
- There are close links between ‘Open Science’ and traditional modeling paradigms!
- By **opening** research, we can examine the certainty of statistical results.
- Does β_1 really equal 7.12452?
- How can we "take the ‘con’ out of econometrics"?

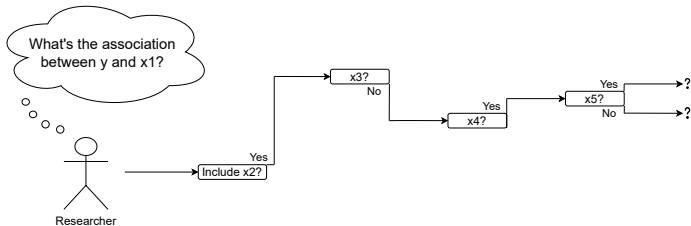
- There are close links between ‘Open Science’ and traditional modeling paradigms!
- By **opening** research, we can examine the certainty of statistical results.
- Does β_1 really equal 7.12452?
- How can we "take the ‘con’ out of econometrics"?
- Transparency and more computational procedures!

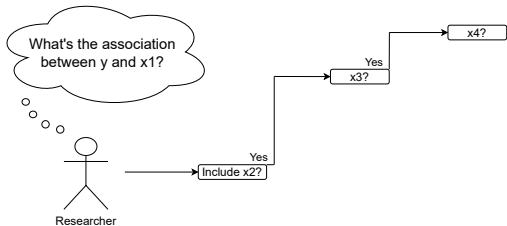


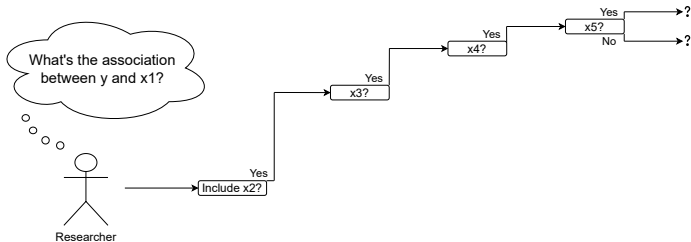


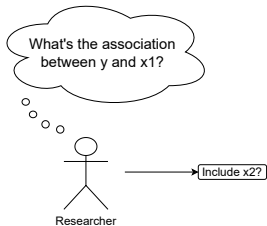


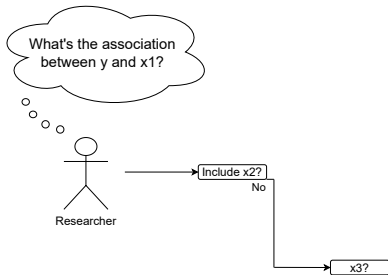


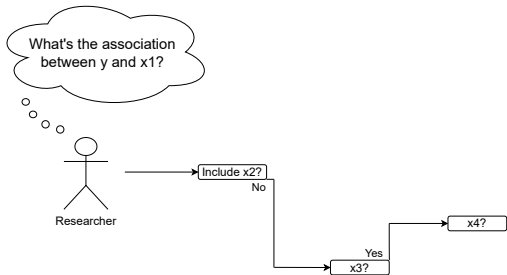


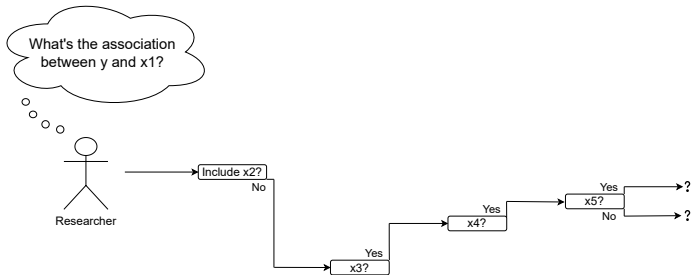






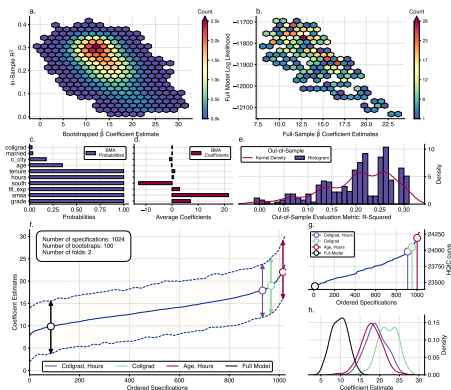






Introducing RobustiPy as a solution!

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



Especially lacking in 'Care' space; reliant on 'classical' approaches!!



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:


Public

[Edit file](#)
[Unwatch](#)

main
1 branch
2 tags

[Add file](#)
[Code](#)

commit	Update union placeholder for github repo header	11e173f · 5 days ago	489 Commits
	Create publish-to-pypi.yml	3 years ago	
	Update union placeholder for github repo header	5 days ago	
	Update union placeholder for github repo header	5 days ago	
	removing cpu default for sim1 example	last month	
	Remove nested loops in _check_colinearity, resolves #409	last week	
	fit test cases for utils.py	last month	
	Take *.pyrib out of gitignore	4 months ago	
	Try not to break the CITATION.cff?	3 months ago	
	Update code of conduct and contributing.md file (as well...	last year	
	Initial commit	5 years ago	
	Add link to paper and citing data into readme.md	3 months ago	
	Update code of conduct and contributing.md file (as well...	last year	
	pushed to pypi	3 months ago	
	Testing readthedocs re-push.	3 months ago	
	Add sphinx_rtd_theme to requirements.txt for readthed...	last year	

[README](#)
[Code of conduct](#)
[Contributing](#)
[GPL-3.0 license](#)



RobustIPy

[PyPI](#)
[conda](#)
[pypi](#)
[3.30](#)
[license](#)
[GPL-3.0](#)
[DOI: 10.5281/zenodo.15750930](#)

Welcome to the home of **RobustIPy**, a Python library for the creation of a more robust and stable model space. RobustIPy does a large number of things, included but not limited to: high dimensional visualisation, Bayesian Model Averaging, bootstrapped resampling, (in- and)out-of-sample model evaluation, model selection via Information Criterion, explainable AI (via [SHAP](#)), and joint inference tests (as per [Simonsohn et al. 2019](#)).

Full documentation is available on [Read the Docs](#). The first release is indexed onto Zenodo [here](#).

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

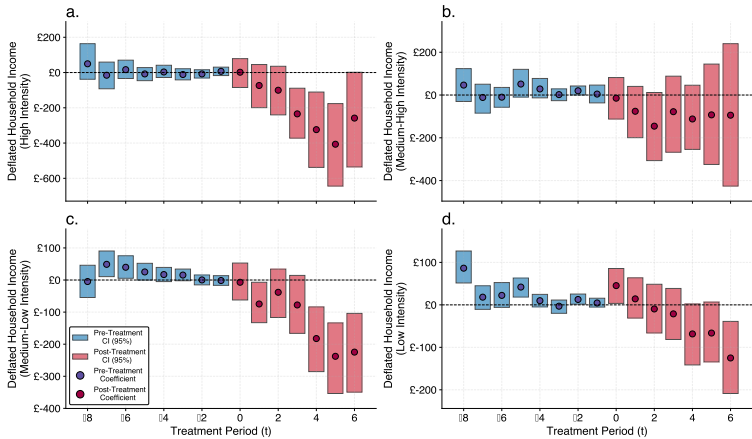
*Denotes joint lead authorship

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.*



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](#)

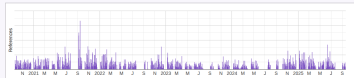
Contributors 3

[dhvalden](#) Daniel Valdenegro Ibarra[crahal](#)[MPetrillo25](#)

Adventures in Open and Assumption-Free Research

[Help using the search](#)

[Results timeline \(accessibility information\)](#)



Search

References (15408)

[View all 15408 results](#) >

SEP
18
2025

Children's Wellbeing and Schools Bill

Baroness Smith of Malvern

18 September 2025

The noble Lord is right that noble Lords have spent considerable hours on this—some more than others, I think. It is for to say, Amendment 595C, tabled by my noble friend Lord Wootton of Innesgowrie, seeks assurances, via a published statement, on law...

Lords Chamber

SEP
18
2025

Children's Wellbeing and Schools Bill
Lord Watson of Invergowrie
18 September 2025

My Lords, if my Amendment 505C appears something of an outlier, it is only because it is. It would have been more appropriate for it to have been introduced in Part 1 of the Bill. But I have to be honest and say that, at the time, I was unaware of it as ...

Lords Chamber

Baroness Smith of Halvern 2

468

← then

The noble Lord is right that noble Lords have spent considerable hours on this—some more than others, I think it is fair to say.

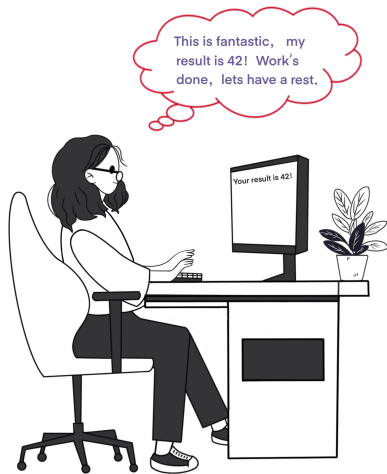
Amendment 529C, tabled by my noble friend Lord Watson of Invergowrie, seeks assurances, via a published statement, on how the Human Rights Act 1998 applies to private providers of publicly funded children's social care services. The amendment seeks a pause in the commencement of Clauses 12 to 18 of the Bill until such a statement is made. These clauses cover measures relating to provider and financial oversight, profit caps, ofsted and Secretary of State powers to impose monetary penalties, and information sharing between ofsted and the Secretary of State.

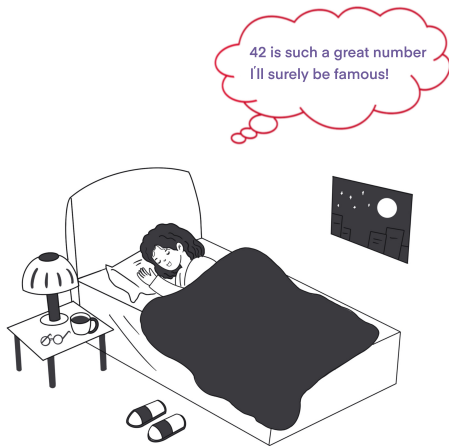
I thank my noble friend for bringing important human rights considerations to our attention. Protecting the human rights of all children is of paramount importance, and the Bill is delivering a number of important measures to stop vulnerable children falling through cracks in services. I completely understand the reasoning behind my noble friend's amendment and would like to provide the Committee with reassurances on the robustness and broad applicability of the human rights legal framework, including for children's social care services.

Section 6 of the Human Rights Act 1998 makes it unlawful for public authorities to act incompatibly with the European Convention on Human Rights, and this duty extends to private providers when they are performing functions of a public nature. There is no statutory definition of "functions of a public nature", but relevant factors considered by the courts include the extent to which the body, in carrying out the relevant function, is publicly funded, exercising statutory powers, taking the place of central government or local authorities, or providing a public service. This broad-scope approach is intentional to ensure that a wide range of services can be captured, as appropriate.

Any statement that the Government make, as proposed by my noble friend, would not be legally binding. Whether a provider is carrying out functions of a public nature and is therefore in scope of the Human Rights Act is for the courts to determine, based on the facts in each specific case. However, to be clear, the Government's position is that, where a court required to decide whether the private provision of care children's **social care** services—services that local authorities are legally required to provide—falls in scope of the Human Rights Act 1998, we believe that the court would conclude that they were indeed performing a function of a public nature and would therefore be in scope.

My noble friend has identified the distinction between the approach taken here and that taken in the Mental Health Bill. As he said, the DHSC is making it explicit in that Bill that certain mental health services are deemed to be exercising a public function for the purposes of the Human Rights Act, but those amendments

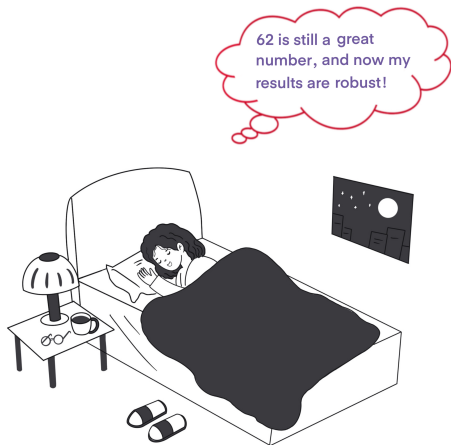


















Pseudo-Random Number Generation (Cont.)

- 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**.

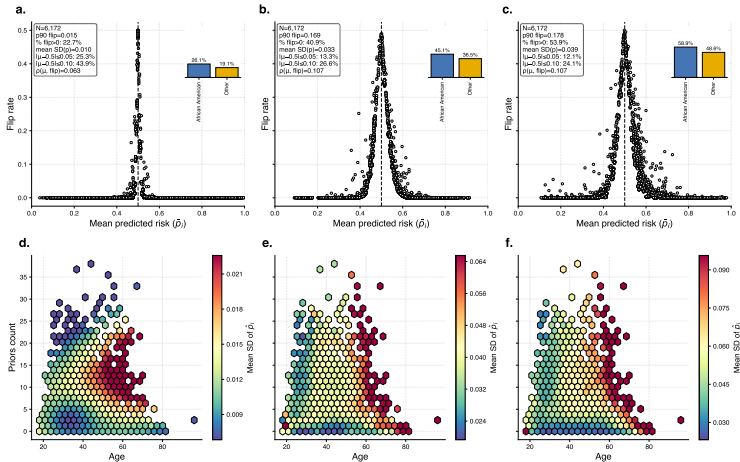
Pseudo-Random Number Generation (Cont.)

- 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 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!