

In the spirit of ‘Open Science’:

crahal.github.io/files/wbs_talk.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

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

Pragmatically ‘Open Science’:

1. Allows the 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 aware of p-hacking, however well intentioned.

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 aware of p-hacking, however well intentioned.
 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 aware of p-hacking, however well intentioned.
 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 aware of p-hacking, however well intentioned.
 3. Critically analyze everyone else's data.
 4. Consider using open-source software.
- Disclaimer: presently a LNL Co-ordinator for UKRN.

A definitional note on ‘Replication’

Replication is a hugely undervalued process. The process quite literally maintains the scientific record. There are four types:

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.

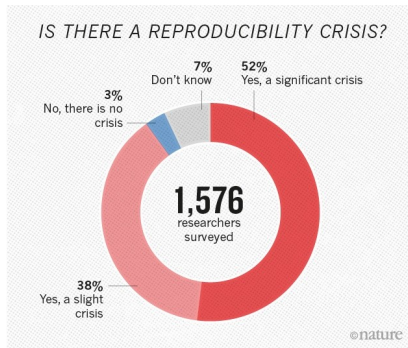
How it started:

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:



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](#) 

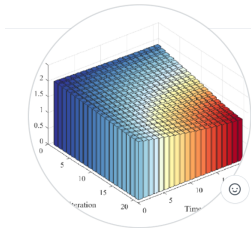
Publishing your data is easy on Harvard Dataverse!

Learn about getting started creating your own dataverse repository here.

[Getting started](#) 

What does 'Open Science' look like?

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

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.

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**!

But what is **semi-structured** data?

“A form of structured data that does not obey the tabular structure of data models associated with relational databases or other forms of data tables, but nonetheless contains tags or other markers to separate semantic elements”

- Different to **tabular** surveys classical social scientists are used to.
- But, often, our goal is to parse it and make it tabular all the same.
 - Semi-structured data found ‘in the wild’ offers huge opportunity.
 - However, it requires a lot to make it ‘research ready’.

Various Open, Reproducible, and Responsible Work

Charlie Rahal

LCDS and Nuffield College, University of Oxford

A talk at the WBS Analytics Insight Series, 14th March, 2024

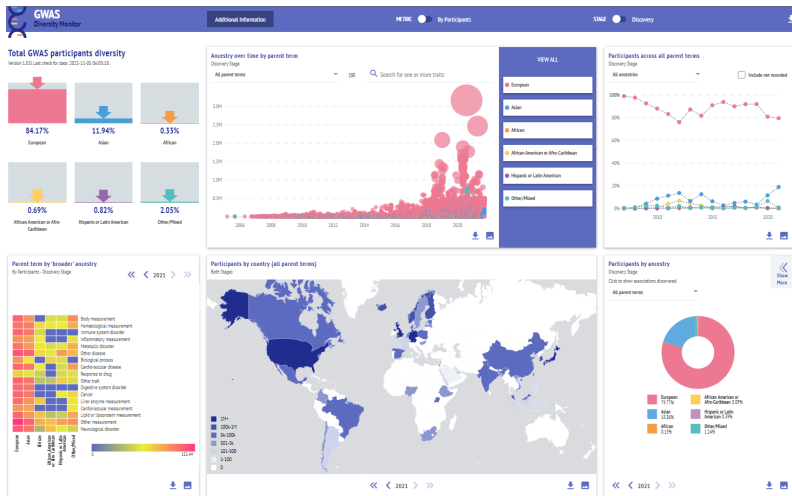


A lot of my semi-structured work is ‘**scientometric**’ and ‘**bibliometric**’ in nature:

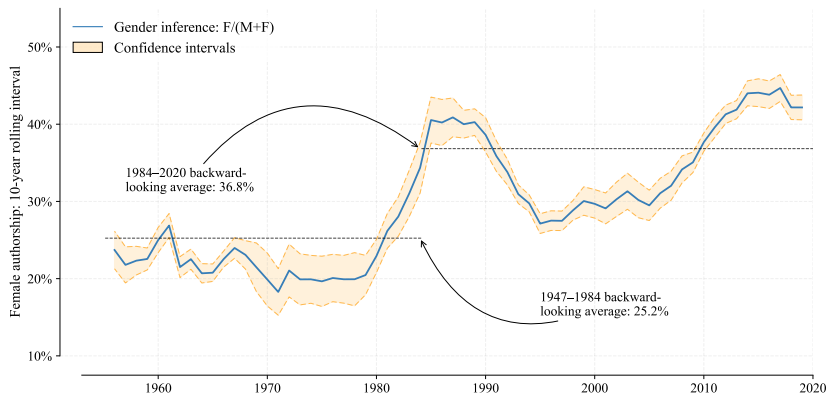
- **Scientometrics** is the field of study which concerns itself with measuring and analysing science.
- **Bibliometrics** is the use of statistical methods to analyze quantitative information pertaining to books, articles and other publications.

These fields are naturally related to ‘**Open Science**’.

The ‘Science of Science’ assess the scientific record.

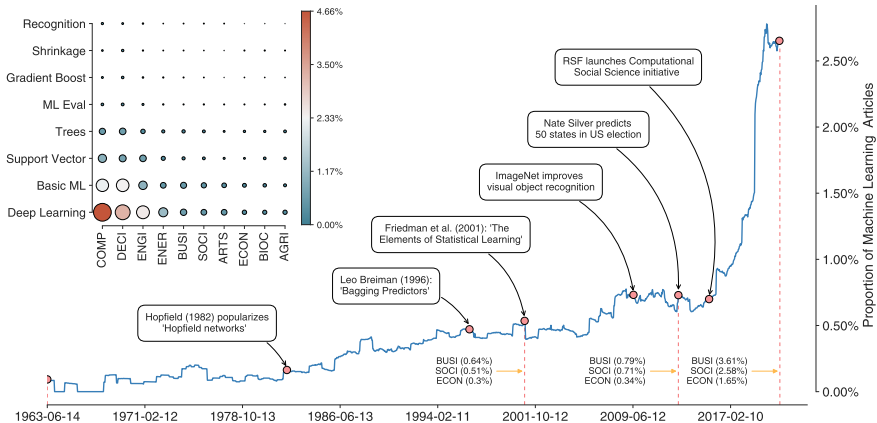


‘The GWAS Diversity Monitor Tracks Diversity in Real Time’
with Melinda C. Mills



‘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)

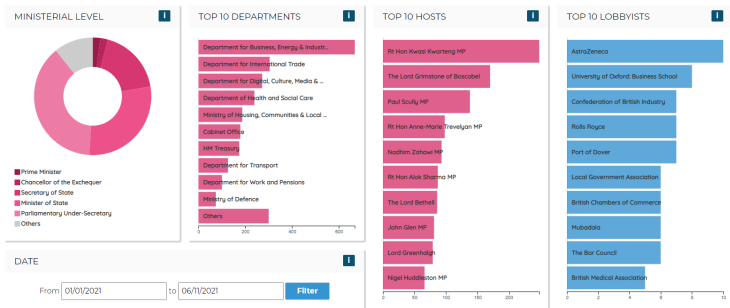
- UK leads the world in terms of ‘Openness’ of government data.
- The value of public government data is increasingly recognised.
 - Increasingly recognised by the ESRC (e.g. ADRNs).
- For better or worse:
 - I’m more interested in ‘hard to use’, semi-structured data.
 - Using curated data from the ONS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

TRANSPARENCY INTERNATIONAL UK fighting corruption worldwide Open Access Home Other versions **DONATE** About

2,623 out of 74,435 meetings SEARCH BY LOBBYIST, HOST, SUBJECT **Reset**  

UK GOVERNMENT

LOBBYIST



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... **into** a .pdf!
(This is a **direct** downgrade in the Berners-Lee 5* schema.)
4. Naturally, of course, there are no supplier UUIDs.

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

A3131

Expenditure Over Threshold

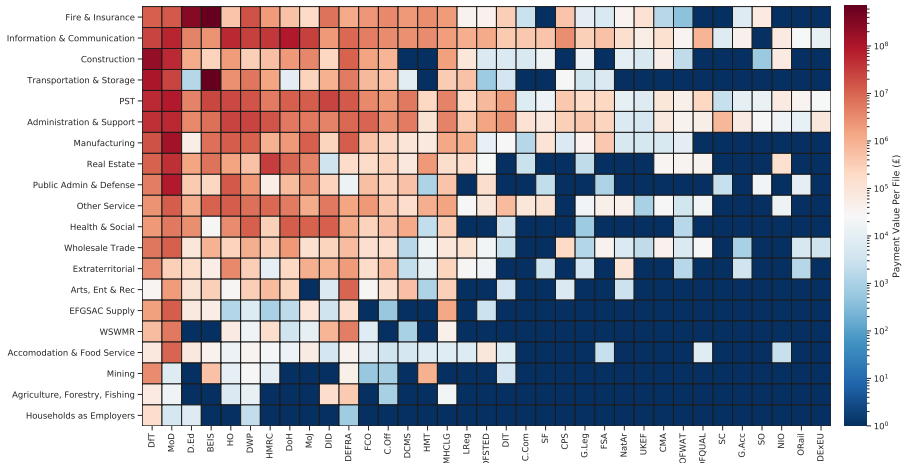
Department/Agency	City	Agency CCG	Agency	Service Name	Signature	Transaction number	AP Amount	VAT registration number	Postcode/Invoice number
Department of health	Beckenham	Beckenham CCG	03042601	Clinical/Medical-Other Public Sector	COMMUNITY SERVICES	11599633	668,000.00		6100071865
Department of health	Beckenham	Beckenham CCG	03042601	Clinical/Medical-Other Public Sector	COMMUNITY SERVICES	11599635	100,000.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Other Gen Supplies & Inv	COMMUNITY SERVICES	11599637	15,666.67		6100071869
Department of health	Beckenham	Beckenham CCG	03042601	Clinical/Medical-Other Public Sector	WYLLIE MASTERS	11599642	64,000.00		6100071867
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11600729	68,226.00	6549438567	610008944
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING - NON ACUTE	11600730	68,226.00		610008944
Department of health	Beckenham	Beckenham CCG	03042601	Clinical/Medical-Other Public Sector	COMMUNITY SERVICES	11615604	30,000.00		6100071863
Department of health	Beckenham	Beckenham CCG	03042601	Clinical/Medical-Independent Sector	ACUTE COMMISSIONING	11615704	38,526.24	787424655	HL02055682
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11615714	58,147.4		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Clinical/Medical-Other Public Sector	HOSPICES	11623337	218,000.00		6100071867
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11634001	4,440.44		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11634002	21,124.14		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11642864	25,672.53		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11642923	-30,000.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11642924	20,000.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11654425	-3,662.11		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	AMBUULANCE SERVICES	11663375	587,757.29	6549431392	61001114
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11666893	768,636.00		61001114
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11668507	760,530.34		61001114
Department of health	Beckenham	Beckenham CCG	03042601	Cont Care Adult 100% Fully Funded	FUNDED NURSING CARE	11669562	26,412.00	2359731641	ALM036320TALM036588
Department of health	Beckenham	Beckenham CCG	03042601	Charges From CSU	RECHARGES	11669599	85,249.00	GB55442045	731106994
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11671003	278,828.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11680115	82,568.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Hilton-Foundation Trusts	COMMUNITY SERVICES	11701702	228,968.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Hilton-Foundation Trusts	COMMUNITY SERVICES	11701703	141,917.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Hilton-Foundation Trusts	COMMUNITY SERVICES	11701710	807,726.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Income Mx < 11%	BALANCE SHEET	11720068	47,733.1	6544623331	1209100565770 MAR 18
Department of health	Beckenham	Beckenham CCG	03042601	Income Mx < 11%-Student Loans	BALANCE SHEET	11720069	738.00	6544623331	1209100565770 MAR 18
Department of health	Beckenham	Beckenham CCG	03042601	National Insurance < 1 vNHS-RMS	BALANCE SHEET	11720065	29,114.84	6544623331	1209100565770 MAR 18
Department of health	Beckenham	Beckenham CCG	03042601	National Insurance < 1 vNHS-RMS	BALANCE SHEET	11720066	19,284.18	6544623331	1209100565770 MAR 18
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11724081	213,245.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11724082	268,141.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11724083	27,124.14		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11722099	266,424.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11724114	224,749.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Clinical/Medical-Other Public Sector	CLINICAL SERVICES	11724123	440.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11739262	-20,056.24		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11739269	53,540.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Hilton-Foundation Trusts	COMMUNITY SERVICES	11748400	249,906.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11748388	47,900.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11771723	10,124.14		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11771733	2,691,912.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Hilton-Foundation Trusts	MENTAL HEALTH CONTRACTS	11777627	25,202.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Clinical/Medical-Independent Sector	COMMUNITY SERVICES - NON ACUTE	11777628	1,000,000.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Hilton-Foundation Trusts	MENTAL HEALTH CONTRACTS	11782426	1,728,242.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Hilton-Foundation Trusts	MENTAL HEALTH CONTRACTS	11801176	36,000.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ESTATES AND FACILITIES	11801415	51,209.79		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Hilton-Foundation Trusts	CHILD AND ADOLESCENT MENTAL HEALTH	11810820	258,781.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Hilton-Foundation Trusts	CHILD AND ADOLESCENT MENTAL HEALTH	11813400	55,506.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Clinical/Medical-Other Public Sector	LEARNING DIFFICULTIES	11813401	25,000.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Clinical/Medical-Other Public Sector	HOSPICES	11815190	276,741.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Clinical/Medical-Other Public Sector	HOSPICES	11815191	276,741.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11824000	1,869,511.00	654226016	6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11824001	250,000.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11854467	250,000.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11854468	250,000.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	COMMUNITY SERVICES	11903191	65,992.39		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	NHS 111	11923323	71,747.88		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Clinical/Medical-Independent Sector	COMMUNITY SERVICES	11923324	26,700.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11923373	1,137,848.99		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11931400	1,455,963.00	654923417	6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11931401	1,137,848.99	654923417	6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11933581	-129,548.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11951210	1,167,569.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11951211	-200,561.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11951212	229,254.29		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	PLANNED CARE	11951290	-411,000.00		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Cont Care Adult 100% Fully Funded	CHC ADT FULL FUND PERMS HLTH PRG	11951291	269,574.97		6100071861
Department of health	Beckenham	Beckenham CCG	03042601	Health Svc Res NHS Trust-Contract Baseline	ACUTE COMMISSIONING	11959239	68,226.00	6549438567	6100071861

A3131

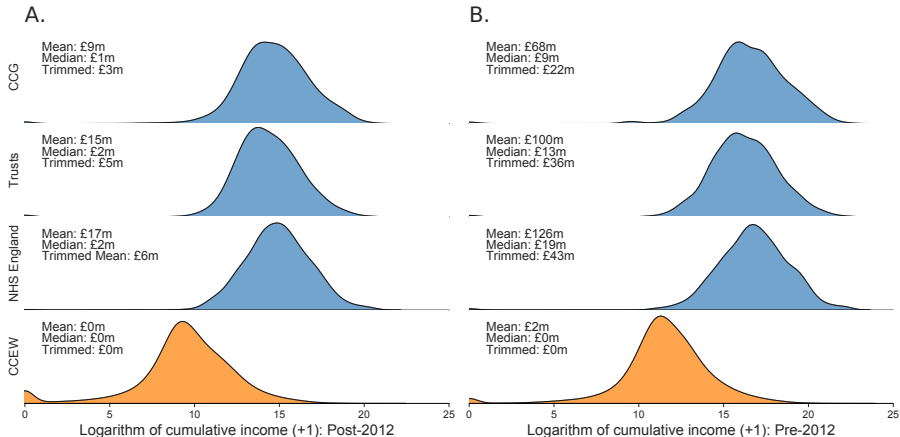
Expenditure Over Threshold

Department/Agency	City	Health CCG	Supply	Contract	Service	Supplier	Transaction number	AP Amount	VAT registration number	Postcode	Invoice number
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Other Public Sector	COMMUNITY SERVICES	BEXLEY COUNCIL	11599653	668,000.00			6100071665
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Other Public Sector	COMMUNITY SERVICES	BEXLEY COUNCIL	11599652	100,000.00			6100071661
Department of health	NHS	Bevery CCG	3054261	Other Gen Supplies & Inv	COMMUNITY SERVICES	BEXLEY COUNCIL	11599651	15,666.99			6100071667
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Other Public Sector	WYLLIE RESTAURANT	BEXLEY COUNCIL	11599652	64,000.00			6100071657
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	UNIVERSITY COLLEGE LONDON HOSPITALS NHS FOUNDATION TRUST	11607291	68,226.00	6544836567		610008644
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	UNIVERSITY COLLEGE LONDON HOSPITALS NHS FOUNDATION TRUST	11607290	68,226.00			610008644
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Other Public Sector	COMMUNITY SERVICES	BEXLEY COUNCIL	11599654	30,000.00			6100071655
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Other Public Sector	COMMUNITY SERVICES	BEXLEY COUNCIL	11599654	30,000.00			6100071655
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Independent Sector	ACUTE COMMISSIONING	NHS HEALTH LTD	11615104	38,528.24	787424655		HL00555682
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	NHS HEALTH LTD	11615103	48,167.74			HL00555682
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Other Public Sector	HOSPICES	GREENWICH & BEXLEY COMMUNITY HOSPICE	11623337	218,000.00			871
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Other Public Sector	ACUTE COMMISSIONING	LEWISHAM & GREENWICH NHS TRUST	11636201	4,440,144.00			84426527
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	NHS PROPERTY SERVICES LTD	11637401	218,124.14			48426526
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	NHS PROPERTY SERVICES LTD	11642864	25,672.53			48413715
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	CARTFORD AND GRAVESEND NHS TRUST	11642862	-308,000.00			2000004184
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	CARTFORD AND GRAVESEND NHS TRUST	11642863	204,342.00			2000004139
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11654425	-3,662,119.34			CM00218970
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11655539	26,332.44			CM00218970
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	AMBUCLANCE SERVICES	LONDON AMBUCLANCE SERVICE NHS TRUST	11665375	587,757.29	654811392		ML011114
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11666893	768,636.00			227489
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11668505	760,530.34			233561
Department of health	NHS	Bevery CCG	3054261	Cont Care Adult 100% Fully Funded	FUNDED NURSING CARE	BURA CARE HOMES	11669562	26,412.00	2359711641	ALM305302YCALM305585	
Department of health	NHS	Bevery CCG	3054261	Charges From CSU	RECHARGES	NHS SOUTH EAST CSU	11669995	85,249.00	GB554442045		731105994
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11671003	278,380.00			1206094
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	MOORFIELD EYE HOSPITAL NHS FOUNDATION TRUST	11680115	82,568.00			CO0126903
Department of health	NHS	Bevery CCG	3054261	Hilton-Foundation Trusts	COMMUNITY SERVICES	OXLEAS NHS FOUNDATION TRUST	11701702	228,948.00			378586
Department of health	NHS	Bevery CCG	3054261	Hilton-Foundation Trusts	COMMUNITY SERVICES	OXLEAS NHS FOUNDATION TRUST	11701701	141,917.00			378532
Department of health	NHS	Bevery CCG	3054261	Hilton-Foundation Trusts	COMMUNITY SERVICES	OXLEAS NHS FOUNDATION TRUST	11701710	807,726.00			378922
Department of health	NHS	Bevery CCG	3054261	Income Mx <11%	BALANCE SHEET	HMC	11720565	47,733.41	654483331		120900556770 MAR 18
Department of health	NHS	Bevery CCG	3054261	Income Mx <11% Student Loans	BALANCE SHEET	HMC	11720566	738.00	654483331		120900556770 MAR 18
Department of health	NHS	Bevery CCG	3054261	National Insurance <1 vNE-RMS	BALANCE SHEET	HMC	11720565	29,114.84	654483331		120900556770 MAR 18
Department of health	NHS	Bevery CCG	3054261	National Insurance <1 vNE	BALANCE SHEET	HMC	11720566	19,284.18	654483331		120900556770 MAR 18
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11724081	213,435.00			2001026
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11724082	268,141.00			20010264
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11724083	27,120.00			20010264
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11722209	266,424.00			231364
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11724114	224,749.00			231364
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Other Public Sector	COMMUNITY SERVICES	LEAF & EVELYN	11734031	440.00			200102676
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Over Under Performance	ACUTE COMMISSIONING	QUEEN VICTORIA HOSPITAL NHS FOUNDATION TRUST	11738262	-20,056.34			1365
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Over Under Performance	ACUTE COMMISSIONING	MOORFIELD EYE HOSPITAL NHS FOUNDATION TRUST	11738269	55,840.00			CO0126948
Department of health	NHS	Bevery CCG	3054261	Hilton-Foundation Trusts	COMMUNITY SERVICES	OXLEAS NHS FOUNDATION TRUST	11748440	249,505.00			17541
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	QUEEN VICTORIA HOSPITAL NHS FOUNDATION TRUST	11748386	47,300.00			201000982
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	LEAF & EVELYN	11771725	182.00			263322
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	CARTFORD AND GRAVESEND NHS TRUST	11771733	2,691,562.00			1000004308
Department of health	NHS	Bevery CCG	3054261	Hilton-Foundation Trusts	MENTAL HEALTH CONTRACTS	SOUTH LONDON AND MAUDSLAY NHS FOUNDATION TRUST	11777627	25,202.00			80002737
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Independent Sector	COMMUNITY SERVICES	BEXLEY COUNCIL	11800436	31,000.00			6100071660
Department of health	NHS	Bevery CCG	3054261	Hilton-Foundation Trusts	MENTAL HEALTH CONTRACTS	OXLEAS NHS FOUNDATION TRUST	11725324	378,917			378917
Department of health	NHS	Bevery CCG	3054261	Hilton-Foundation Trusts	MENTAL HEALTH CONTRACTS	OXLEAS NHS FOUNDATION TRUST	11801176	36,000.00			378908
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	PROPERITY SERVICES LTD	11801175	278,380.00			484138351
Department of health	NHS	Bevery CCG	3054261	Hilton-Foundation Trusts	CHILD AND ADOLESCENT MENTAL HEALTH	OXLEAS NHS FOUNDATION TRUST	11818020	258,781.00			378905
Department of health	NHS	Bevery CCG	3054261	Hilton-Foundation Trusts	CHILD AND ADOLESCENT MENTAL HEALTH	OXLEAS NHS FOUNDATION TRUST	11818040	55,856.00			378904
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Other Public Sector	HEALTHY SUPPLIES LTD	OXLEAS NHS FOUNDATION TRUST	11818100	1,000.00			10000030263
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Other Public Sector	DAY THERAPIES	GREENWICH & BEXLEY COMMUNITY HOSPICE	11818190	276,741.00			589
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Other Public Sector	DAY THERAPIES	DAY THERAPIES INTERNATIONAL	11818585	25,000.00	852426016		610007165877
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Other Public Sector	ACUTE COMMISSIONING	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11824000	1,960,511.00			243430
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Other Public Sector	LEARNING DIFFICULTIES	BEXLEY COUNCIL	11854467	250,000.00			61000717247
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	UNIVERSITY COLLEGE LONDON HOSPITALS NHS FOUNDATION TRUST	11861126	1,167,126			CM00218970
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Independent Sector	COMMUNITY SERVICES	BELLEROVE ROAD SURGERY	11963191	65,592.39			BA-0330156
Department of health	NHS	Bevery CCG	3054261	Healthcare Svc Res CCG	NHS 111	NHS BROMLEY CCG	11963233	71,747.88			10002101149
Department of health	NHS	Bevery CCG	3054261	Clinical/Medical-Independent Sector	COMMUNITY SERVICES	CARFORD TOWN SURGERY	11963234	26,700.00			CM00218970
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11963237	1,137,848.98			234497
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	GUY'S & ST THOMAS HOSPITAL NHS FOUNDATION TRUST	11963140	1,455,963.00	654493417		2106559
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	LEWISHAM & GREENWICH NHS TRUST	11963141	3,133,400.00	654493417		8446655
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	LEWISHAM & GREENWICH NHS TRUST	11963181	-129,548.00			8446655
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11963182	1,167,569.00			233736
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11963183	-200,156.00			CM00218970
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11963187	229,254.29			233542
Department of health	NHS	Bevery CCG	3054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11963190	-411,000.00			CM00218970
Department of health	NHS	Bevery CCG	3054261	Cont Care Adult 100% Fully Funded	CHC AD FULL FUND PERKS HLTH RD	BEXLEY COUNCIL	11963192	269,574.97			6100071665
Department of health	NHS	Bevery CCG	18	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	UNIVERSITY COLLEGE LONDON HOSPITALS NHS FOUNDATION TRUST	11963235	68,226.00	6544836567		610009610

28 / 58

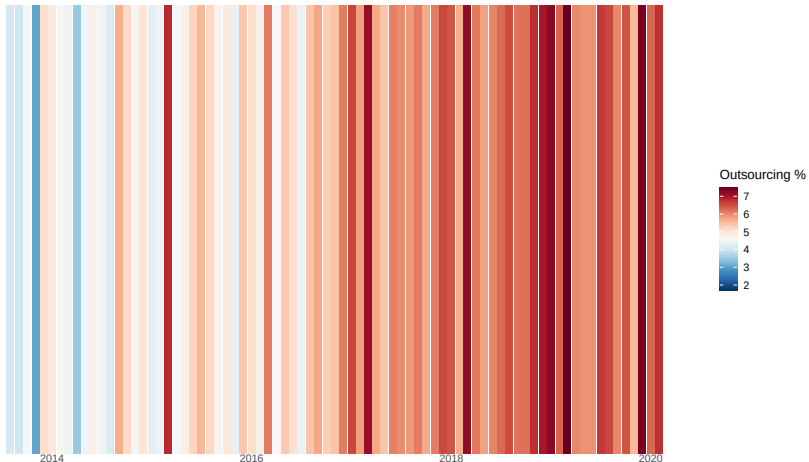


‘Tools for Transparency in Central Government Spending’



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

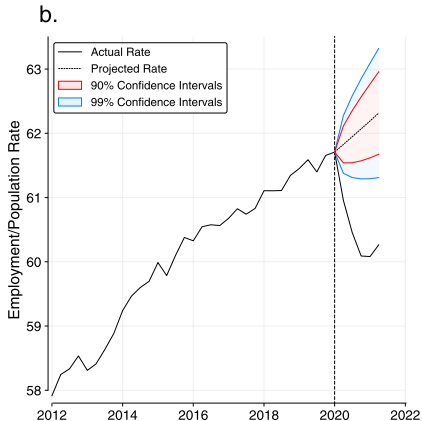
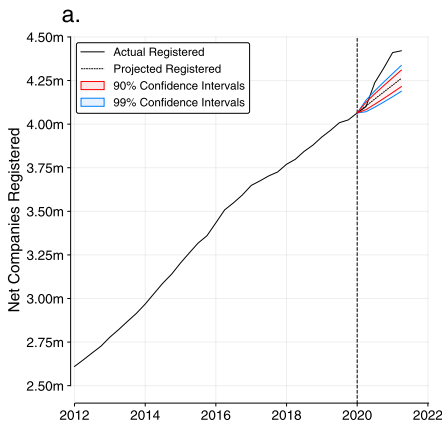
Creeping Privatisation in England's NHS



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

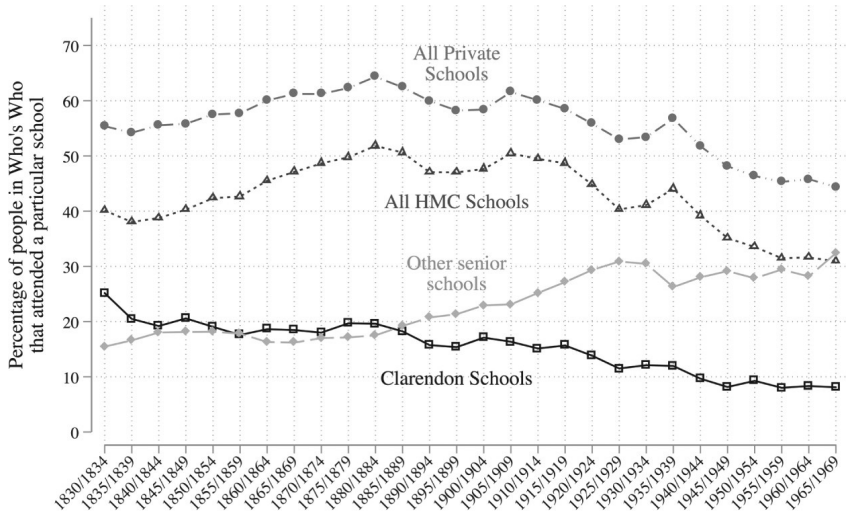


‘Modelling Labour Market Inequalities and the Covid-19 Pandemic’

With Aaron Reeves and Naomi Muggleton

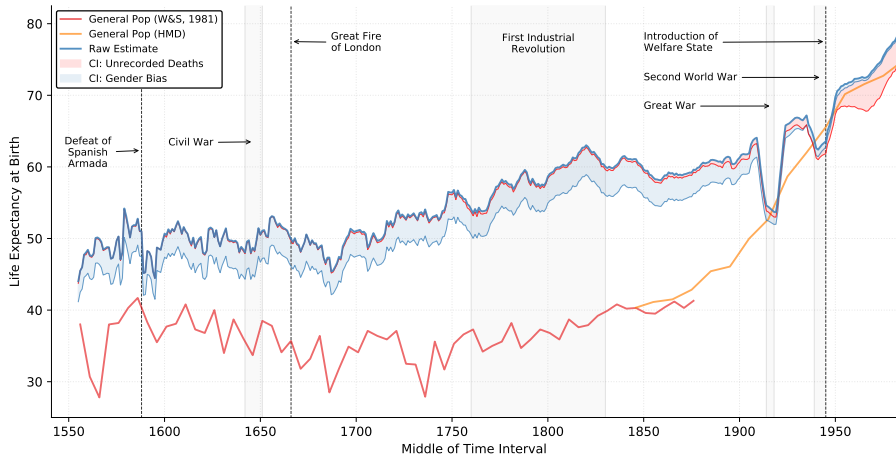
Semi-Structured Data for Inequality Research

- Data doesn't always come from APIs.
- Sometimes, it can come from raw `.html` itself.
- One example of this is when we scraped Who's Who to get information on education background of 'elites'.
- Another is when we scraped `thepeerage.com` to get data on dates of birth and death of British Peers.
- We might want to (there are competing views):
 - Check ToS and `robots.txt`.
 - Contact the webmaster first.



‘The decline and persistence of the old boy’

With Aaron Reeves, Sam Friedman. and Magne Flemmen



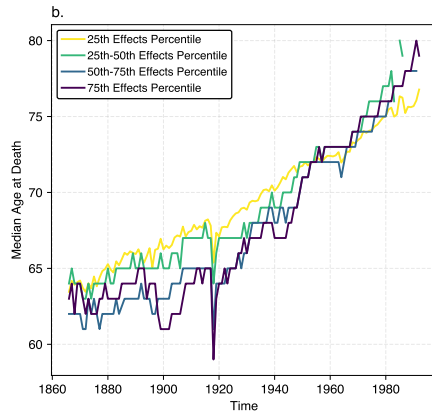
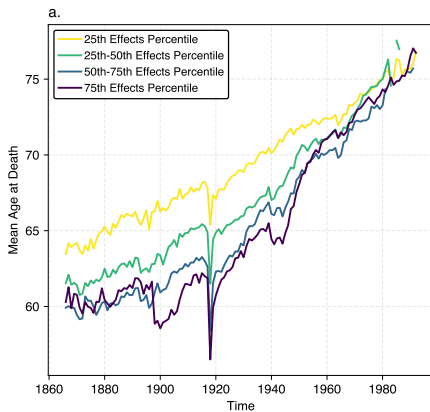
‘The Legacy of Longevity: Inequalities in UK Life Expectancy 1500-2016’ with Aaron Reeves, Felix Tropsch and Darryl Lundy

Ongoing research (led by the wonderful Naomi Mugglestone!) uses probate records from 1858-1996.

The idea is to:

1. Scrape 800k .pdfs containing 24m probates: **Done!**
2. Preprocess images ready for OCR: **Done!**
3. Run OCR: **Done!**
4. Merge with geneological data: **Done!**
5. EDA/lifetables to calculate e_0 by wealth at death...: **Pending!**

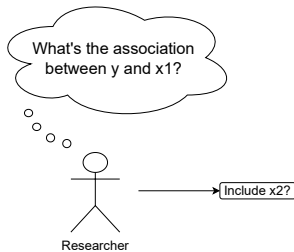
689	ABBOTT Henry Bernard of 22 Inter- Middleton died 18 April 1901 at House, Liverpool Street 1901 Harrington Abbott widow. Effects £221 12s. 6d.	brood Street Green James Connelton was 8 May to Louise Harrington Abbott widow. Effects £16 6s. 6d.
10	ABBOTT Isaac of Thrapston Northamptonshire bootmaker died 16 October 1908 Administration £1 December to Sarah Hannah Abbott widow. Effects £221 12s. 6d.	10 ABBOTT Susan Georgina of Barling House Grange- Eastbourne (wife of Henry Susan Abbott) died 11 1908 Probate London 19 July to the said Henry Susan Abbott. Effects £101 12s. 6d.
12	ABBOTT James of The Elms Home Grove-road Ramsgate licensed victualler died 1 February 1908 Probate January 1908 to Elizabeth Abbott widow. Effects £284 12s. 6d.	12 ABBOTT Thomas of Hincaster gentleman died 17 Novem- ber 1907 Probate London 9 July to William John Stirling and Henry David Boddy as executors. Effects £268 12s. 6d. Residue £7109 14s. 2d.
86	ABBOTT John of 13 Lennard-road Leytonstone Essex died 29 January 1908 Administration London 3 March to John Benjamin Abbott contractor. Effects £449 17s. 2d.	86 ABBOTT Thomas of 103 Abbeville-road Rochdale-road Ma- chester died 10 March 1908 Probate Manchester 29 Apr to Thomas Edward Miller and William Sells compositors. Effects £108 11s.
103	ABBOTT John of Hatched Lanehouse died 20 February 1908 Administration London 25 February to Thomas Abbott farmer. Effects £298 4s.	103 ABBOTT Thomas of 30 Marshall-street Middlesbrough died 7 June 1908 Administration London 9 July to Ede Abbott and George Hedges Abbott executors. Effects £559 16s. 12d.
103	ABBOTT John of Mayfield House Northfield House Yorkshire died 28 June 1908 Probate London 10 September to Mary Abbott widow and William Abbott railway clerk. Effects £216 12s. 6d.	103 ABBOTT William of Kiveton-road Rawmarsh Rotherham Yorkshire colliery-undertaker died 1 November 190 Probate Wakefield 12 February to Ellen Abbott widow Effects £134 10s. 8d.
103	ABBOTT John White of 4 Westbury-gardens Hanger Lane died 21 April 1908 Probate London 27 April to Gertrude White Abbott widow. Effects £213 12s. 6d.	103 ABBOTT William of Mercatula Lincolnshire died 23 Janua 1908 Probate Lincoln 23 March to John Elmy Chatterto clerk to the Governor. Effects £213.
103	ABBOTT Joseph of 8 Progression-roads Clapham Field Barnetville Lincolnshire engraver died 7 November 1908 Administration London 9 December to Maria Jane Abbott widow. Effects £432 14s. 2d.	103 ABBOTT William of 61 Post-street Finsbury Wagon maker died 14 October 1908 Administration Liverpo 10 November to Sarah Ann Lowe (wife of Edwin Lowe Effects £17 3s.
103	ABBOTT Lucia Augusta of Banastone Levensand Islands gentle died 8 June 1908 Administration 19th W United London died 14 January to Paget Augusta Wade mar- ried attorney of Cardiff. Effects £101 12s. 6d.	103 ABBOTT William of 31 York-road Southend-on-Se Essex died 14 April 1908 Administration South W London 3 July to Frank Abbott commercial traveller. Effects £50 10s. 2d.
103	ABBOTT Mary of Colchester Lincolnshire widow died 13 October 1908 Probate London 14 December to Charles Alfred Ernest Abbott bank clerk and George Edward Abbott bank manager. Effects £1042 15s. 4d.	103 ABBOTT Henry of The Cranhill Cottage near Woodrove Derbyshire died 18 May 1908 Probate London 3 December to Elizabeth Abbott widow. Effects £201 7s. 12d.
103	ABBOTT Mary of 11 Ashfield-road Park Village Georgina Abbott died 1 July 1908 Administration London 1 August to the said George Abbott gardener. Effects £449 12s. 10d.	103 ABBOTT Jane of West Beckham Norfolk widow died 27 Apr 1908 Probate Norwich 10 June to James Kemp maid Effects £109.
103	ABBOTT Percy William of 35 Ashfield Park Villa court Park Mansfield died 5 May 1908 at sea. Admini- stration London 3 August to Arthur Abbott chemist's manager. Effects £241 12s. 6d.	103 ABBOTT Robert of Barton Norfolk retired farmer died 4 Octob 1908 Probate Norwich 17 October to Robert James Ab Effects £284 12s.
103	ABBOTT Samuel of Cornwell Yorkshire died 21 June 1908 Administration (with Wm York 8 December to George Albert York brother and Christopher Arthur Rad millers. Effects £43.	103 ABBOTT Daphne Mary Thelma of 28a Rue de Malmaison Engli sh Borne 11 January (wife of Sir William Neville And herbert) died 14 September 1908 Administration Lond 29 October to George Herbert Caine commercial dire Effects £210.
103	ABBOTT Samuel Mitchell of Elsholm Fife-shire-road Brein Glasgow Bromley Kent died 24 August 1908 Admini- stration London 11 December to Elizabeth Abbott widow. Effects £112.	103 ABBOTT Anne of Upington Stasley-road Sutton Surrey wid died 31 October 1908 Administration London 1 Decem to Mary Anne Stasley widow (wife of Ernest Ben Miller). Effects £268 15s. 10d.
52	ABBOTT Sarah of Chawton Somersetshire spinner died 10 December 1908 Administration (with Wm Leande 23 April to Emily Gowler widow. Effects £53 3s. 10d.	52 ABBOTT Anna Maria of Trench Norfolk widow died 12 Ma 1908 Probate London 3 June to the executor Wili Frederick Kemp clerk. Effects £467 12s. 6d.
69	ABBOTT Sarah of The Square Bedford Nottinghamshire (wife of Thomas Abbott) died 17 June 1908 Probate 20 July to Thomas Abbott engineer. Effects £39 16s. 6d.	69 ABBOTT Emma of 11 Denestable-road Luton Bedfordshire gen died 18 September 1908 Probate Northampton November to Edna South sister and Alfred William S baker. Effects £2152 16s. 6d.

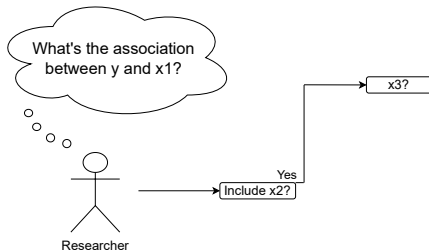


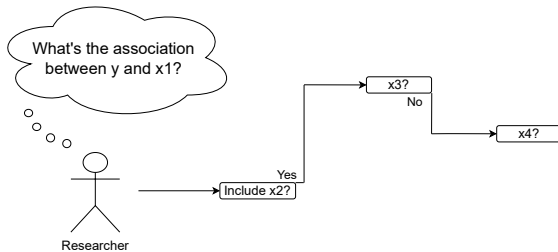
‘Mortality Differentials by Income’ (title TBC!)

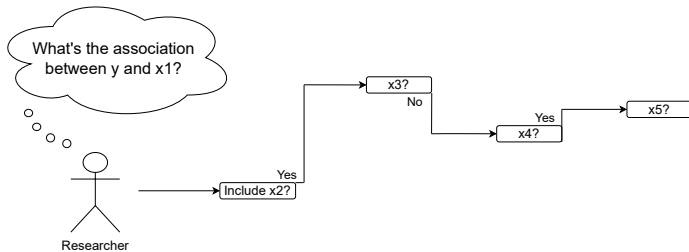
with Naomi Muggleton and Aaron Reeves

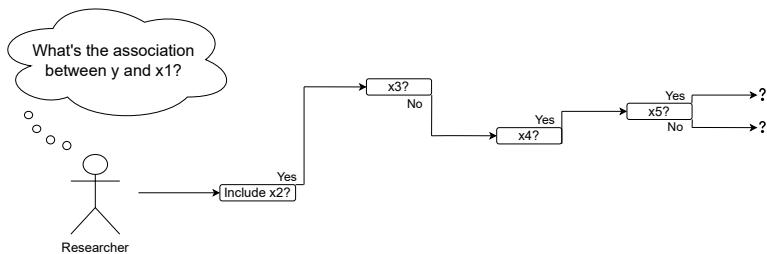
- There are close links between ‘Open Science’ and traditional modeling paradigms!
- By **opening** research, we can examine the certainty of statistical results.
- Does β_1 really equal 7.12452?
- How can we "take the ‘con’ out of econometrics"?

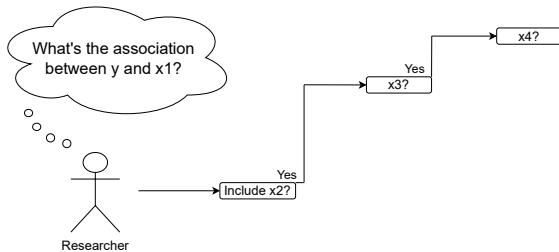


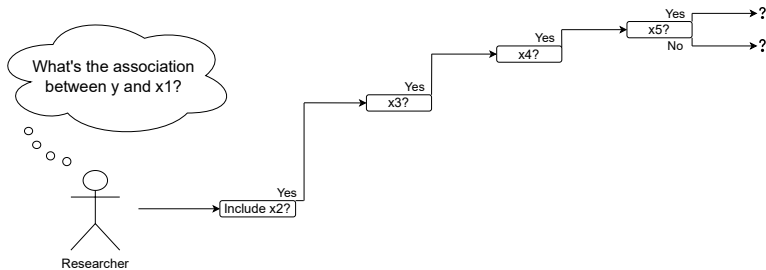


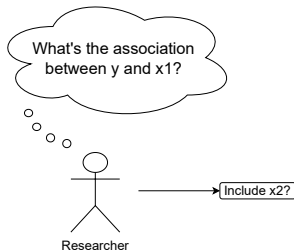


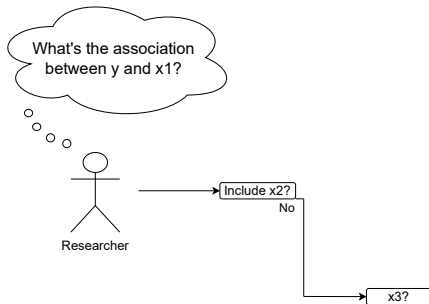


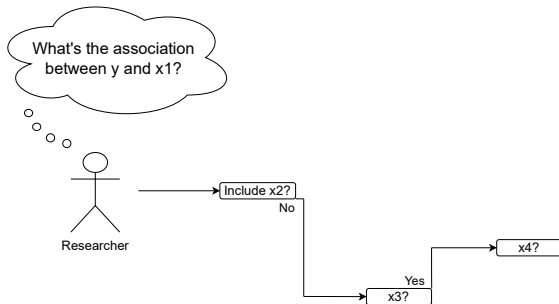


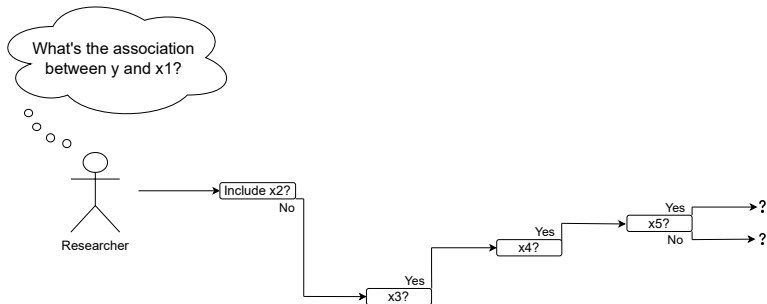


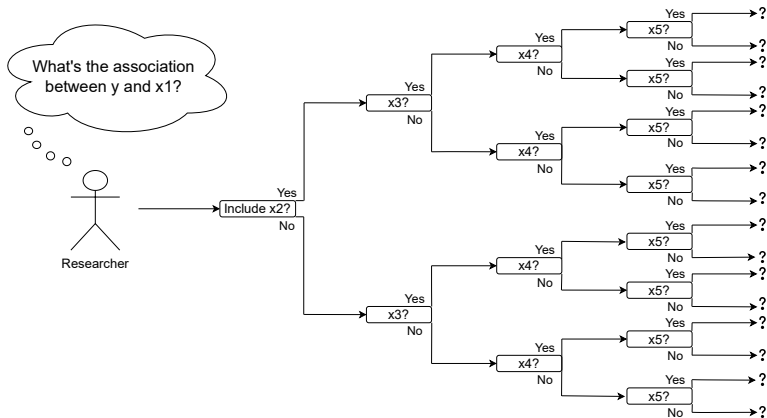




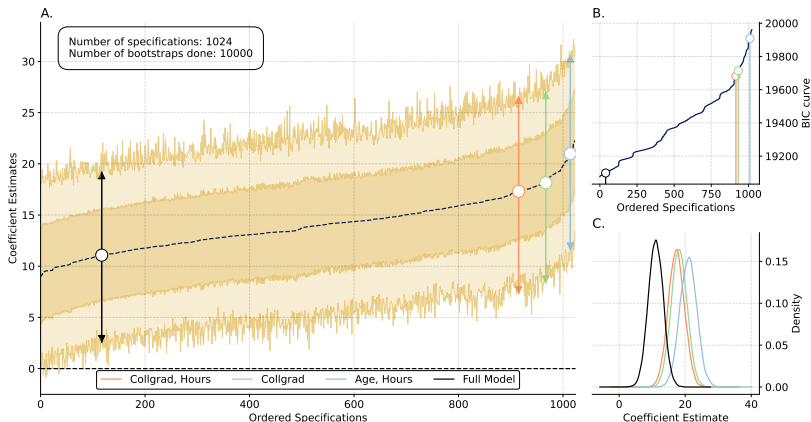




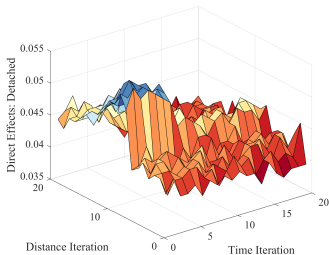
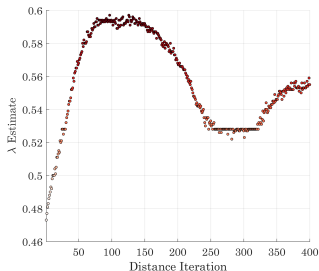
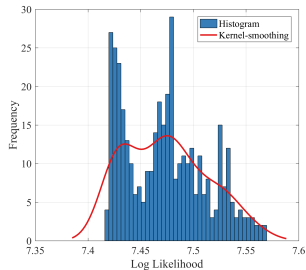
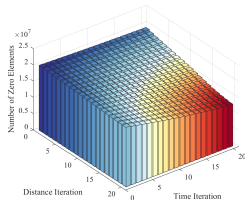




- Our work tackling this type of uncertainty is still in progress!



- The idea is to advance the **specification curve literature**.
- In particular: selection, **averaging**, and **influence analysis**.
- See github.com/centre-for-care/nrobust for updates!



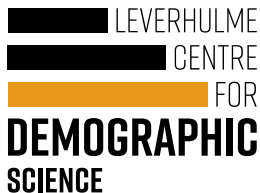
'A Grid Based Approach to Analysing Spatial Weighting Matrix Specification'

On the Responsible use of Pseudo-Random Number Generators in Scientific Research

Charlie Rahal

LCDS and Nuffield College, University of Oxford

A talk at the WBS Analytics Insight Series, 14th March, 2024



What is a PRNG?



ChatGPT

A pseudo-random number generator (PRNG) is a deterministic algorithm creating seemingly random number sequences using a specified seed. Unlike true random generators relying on unpredictable physical processes, PRNGs offer repeatability. This trait is beneficial in research, ensuring consistent results with the same seed, essential for reproducibility in various applications.



“Gott würfelt nicht.”

Source: Einstein on Heisenberg’s indeterminacy principle.

Pseudo-Random Number Generation

An invisible source of uncertainty in the scientific record: PRNGs!

Pseudo-Random Number Generation

An invisible source of uncertainty in the scientific record: PRNGs!

- **Q:** Has anyone ran a program twice, with different results?

Pseudo-Random Number Generation

An invisible source of uncertainty in the scientific record: PRNGs!

- **Q:** Has anyone ran a program twice, with different results?
- **Q:** Has anyone here ever set a '**seed**'? Which seed?

Pseudo-Random Number Generation

An invisible source of uncertainty in the scientific record: PRNGs!

- **Q:** Has anyone ran a program twice, with different results?
- **Q:** Has anyone here ever set a '**seed**'? Which seed?
- Maybe to eliminate variation in algorithms with PRNGs?
 - This seems to be the current 'best practice' advice.

Pseudo-Random Number Generation

An invisible source of uncertainty in the scientific record: PRNGs!

- **Q:** Has anyone ran a program twice, with different results?
- **Q:** Has anyone here ever set a '**seed**'? Which seed?
- Maybe to eliminate variation in algorithms with PRNGs?
 - This seems to be the current 'best practice' advice.
- We argue this is the **opposite of what we want!**
 - We propose you pre-specify **multiple** (complex) seeds.

Pseudo-Random Number Generation

An invisible source of uncertainty in the scientific record: PRNGs!

- **Q:** Has anyone ran a program twice, with different results?
- **Q:** Has anyone here ever set a '**seed**'? Which seed?
- Maybe to eliminate variation in algorithms with PRNGs?
 - This seems to be the current 'best practice' advice.
- We argue this is the **opposite of what we want!**
 - We propose you pre-specify **multiple** (complex) seeds.
 - Or, alternatively, use a subset of the seeds **we** provide.

Pseudo-Random Number Generation (Cont.)

- We need to assess possible variation independent of arbitrary seed choice.
- An extremely important and scarcely researched problem.
- Pseudo-Random Number Generators occur **everywhere!**

Pseudo-Random Number Generation (Cont.)

- We need to assess possible variation independent of arbitrary seed choice.
- An extremely important and scarcely researched problem.
- Pseudo-Random Number Generators occur **everywhere!**
- The variation in estimand can be **huge** (as we'll show).
- We bring attention to this through multiple types of replications.

Pseudo-Random Number Generation (Cont.)

- We need to assess possible variation independent of arbitrary seed choice.
- An extremely important and scarcely researched problem.
- Pseudo-Random Number Generators occur **everywhere!**
- The variation in estimand can be **huge** (as we'll show).
- We bring attention to this through multiple types of replications.
 - Simulations, machine learning, NLP, and inferential research.
- Seed variation at times: i.) can be eliminated, ii.) is not a concern, iii.) cannot be eliminated (requires examination).

The General Premise

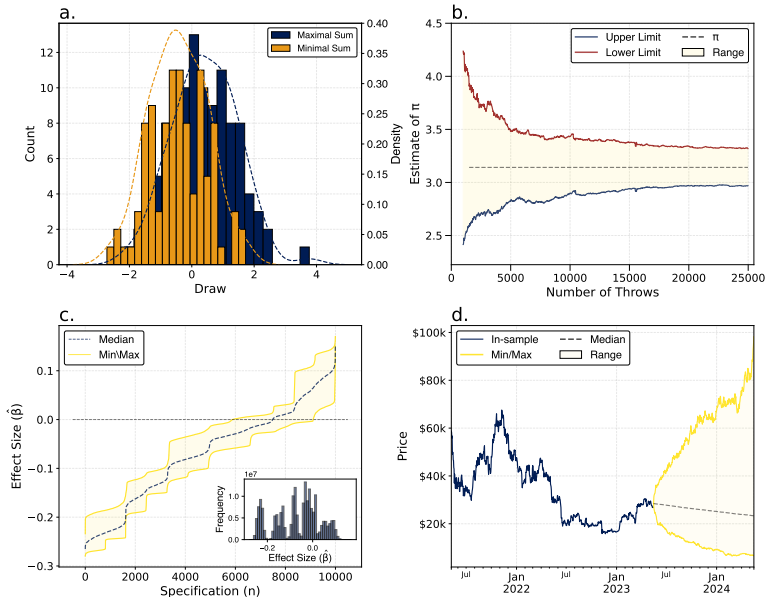
- **Problem Statement:** By setting **one** seed, we ignore the variation of our estimand as a function of how PRNGs were generated: computationally un-intensive, but scientifically dour.

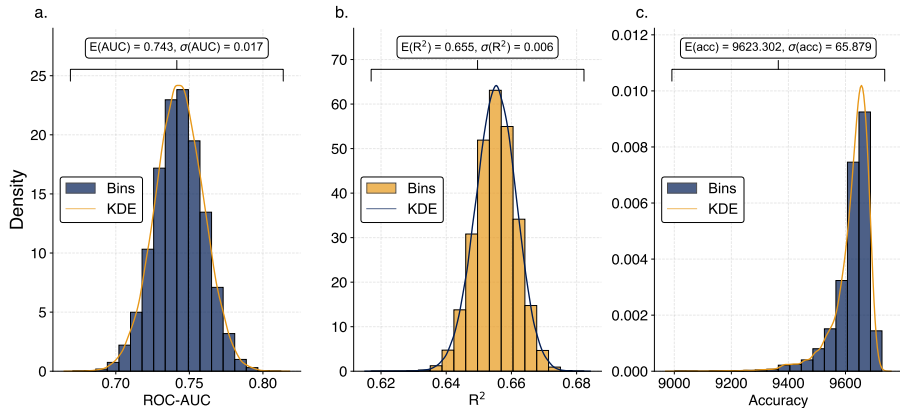
The General Premise

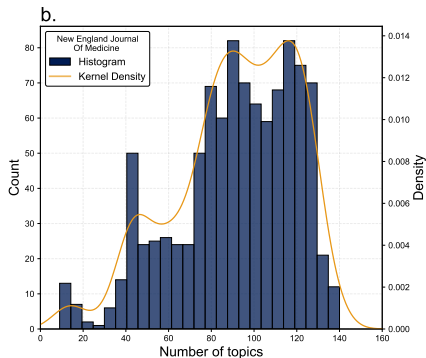
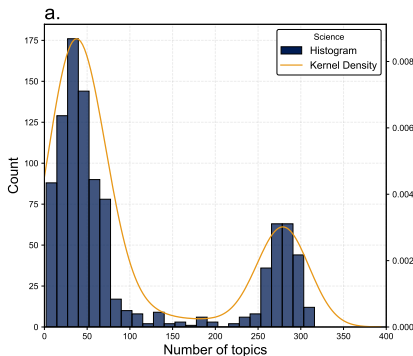
- **Problem Statement:** By setting **one** seed, we ignore the variation of our estimand as a function of how PRNGs were generated: computationally un-intensive, but scientifically dour.
- **Solution:** Visualize the outcome space of a **large number** (10k? 100k?) of seeds simultaneously. This is computationally intensive, but scientifically faithful.

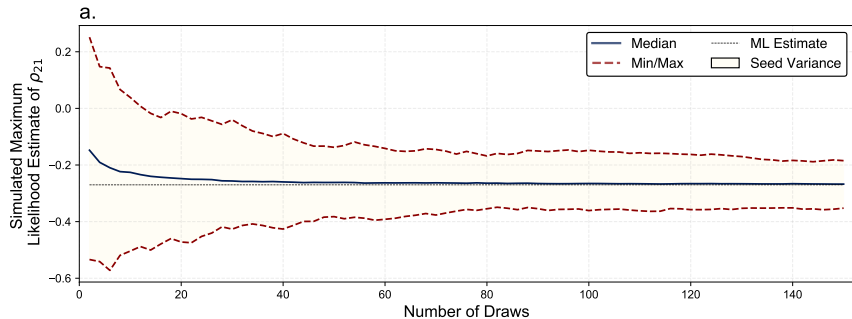
The General Premise

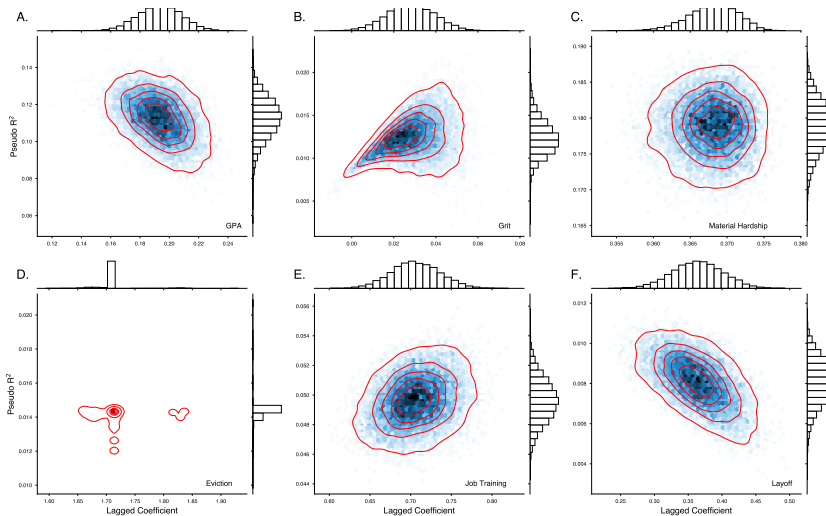
- **Problem Statement:** By setting **one** seed, we ignore the variation of our estimand as a function of how PRNGs were generated: computationally un-intensive, but scientifically dour.
- **Solution:** Visualize the outcome space of a **large number** (10k? 100k?) of seeds simultaneously. This is computationally intensive, but scientifically faithful.
- **Replication:** Similar to specification curves and p-hacking, we can consider whether an original result is in the tail/IQR of the distribution of possible outcome space.









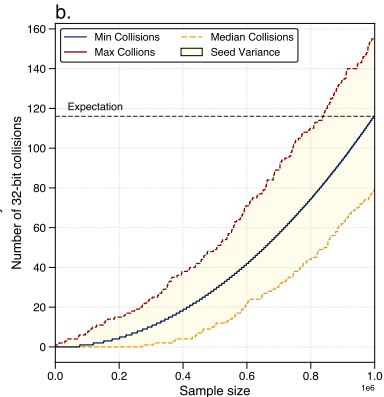
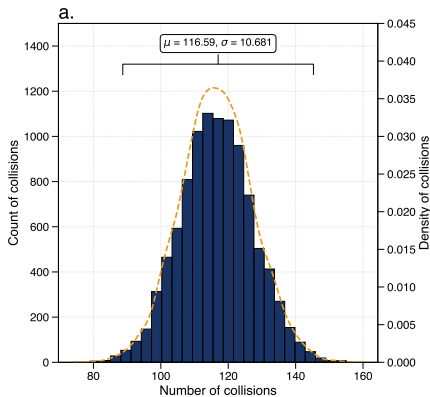


To make things more complicated...

- Hofert (2020): in *some* scenarios, **collisions** exist.
- Collisions: generation of duplicates b/c of insufficient complexity.
- A cherished colleague recently stated to me:

“We wanted to run it [the model] with 100k seeds, but in reality, we only reported 99,905 draws or something, because we were seeing that our results weren’t unique when they should have been. . .”

- This *only* occurs in 32-bit Mersenne Twister implementations.
 - With 32-bit MT: $E(\text{collisions})$ is 116.
 - With 64-bit MT: $E(\text{collisions})$ is $2.71e-08$.
- The default implementation of MT in R is 32-bit. . .



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

- 'Assumption free' social science.
- 'Semi-structured' data found 'in the wild'.
- Openness and Transparency in research.

If a single more robust paper, or even just a slightly more reliable single estimate enters scientific record, everything will be worth while 😊.

Thanks for attending, and a special
thanks to my collaborators!

Don't hesitate to get in touch at:

charles.rahall@demography.ox.ac.uk

and

github.com/crahall!

Temporary page!

$\text{\LaTeX}$ was unable to guess the total number of pages correctly, so if there was some unprocessed data that should have been at the end of the final page this extra page has been added to receive it. If you rerun the document (without altering it) this surplus page will go away, because $\text{\LaTeX}$ now knows how many pages to expect at the end of this document.