

Adventures in Open and Assumption-Free Research

Charlie Rahal

LCDS and Nuffield College, University of Oxford

A talk at the CFC Summer School, 20th September 2024



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. “What’s one piece of scientific knowledge you think everyone should have access to?”

Thought Experiments

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

Thought Experiments

1. “What’s one piece of scientific knowledge you think everyone should have access to?”
2. ‘What’s one scientific myth or misconception you wish more people understood correctly?’
3. ‘What’s an example of a global problem that could be solved faster with open data?’

Thought Experiments (Cont.)

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

Thought Experiments (Cont.)

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

Thought Experiments (Cont.)

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

Thought Experiments (Cont.)

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

Thought Experiments (Cont.)

8. 'If you could instantly make one field of research fully transparent and open, which would it be?'

Thought Experiments (Cont.)

8. 'If you could instantly make one field of research fully transparent and open, which would it be?'
9. 'What's one everyday thing you wish more people understood the science behind?'

Thought Experiments (Cont.)

8. 'If you could instantly make one field of research fully transparent and open, which would it be?'
9. 'What's one everyday thing you wish more people understood the science behind?'
10. 'How can we encourage more people to participate in or contribute to open science?'

Thought Experiments (Cont.)

8. 'If you could instantly make one field of research fully transparent and open, which would it be?'
9. 'What's one everyday thing you wish more people understood the science behind?'
10. 'How can we encourage more people to participate in or contribute to open science?'
11. 'If you could crowd-source one big research project, what would you ask people to help with?'

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.
- As appropriate, discuss Open (CfC) Care Research.

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)
- L^AT_EX (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.

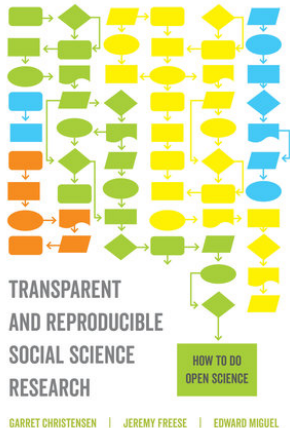
What motivates a talk about ‘Open Science’?

- 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.

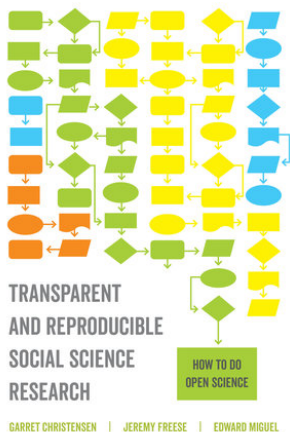
More traditional motivation, just in case you aren't sold...:



This book...

Is most definitely worth
reading!*

More traditional motivation, just in case you aren't sold...:



This book...

Is most definitely worth
reading!*

*Not just because one of
the authors is a coauthor of
a paper included later!

It does an **excellent** job of
motivating responsible re-
search.

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

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

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

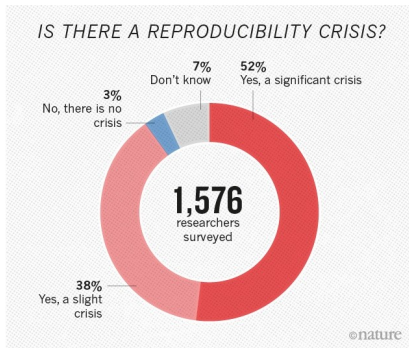
How it started:

Why Most Published Research Findings Are False

John P. A. Ioannidis

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

How it's going:



‘But I do qualitative work!’

How do these ideas relate to researchers doing qualitative work?

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

‘But I do qualitative work! (Cont.)’

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

What does ‘Open Science’ look like?

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



What does ‘Open Science’ look like?

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


HARVARD
Dataverse

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

Deposit and share your data. Get academic credit.

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

[Add a dataset](#)

Organize datasets and gather metrics in your own repository.

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

[Add a dataverse](#)

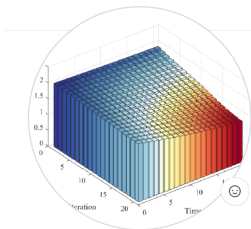
Publishing your data is easy on Harvard Dataverse!

Learn about getting started creating your own dataverse repository here.

[Getting started](#)

What does ‘Open Science’ look like in practice?

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



Pinned

Customize your pins

[GWASReview](#)

Public ::

A Scientometric Review of Genome Wide Association Studies

● HTML ☆ 15 📄 7

[NHSSpend](#)

Public ::

A home for the NHSSpend Library

● HTML ☆ 1 📄 1

[InterModel_Vigorish](#)

Public ::

Introducing the InterModel Vigorish (IMV)

● Jupyter Notebook ☆ 1

[OxfordDemSci/gwasdiversitymonitor](#)

Public ::

Source code for the GWAS Diversity Monitor

● CSS ☆ 12 📄 1

These slides are made with an open source typesetting language!

I_make_this_slide.tex

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

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

guess_what_this_does.py

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

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

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

Thats right: Cats!



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

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

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

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

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

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

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

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

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

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

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

5 Reasons Why Open Science Benefits You

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

5 Reasons Why Open Science isn't Costless

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

Things that Open Science is Not

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

A lot of my work is ‘**scientometric**’/‘**bibliometric**’ in nature.

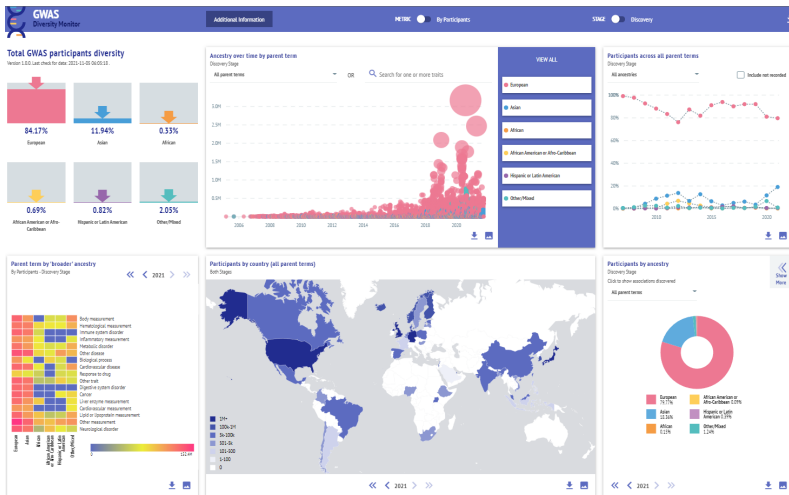
- **Scientometrics** is the field of study which concerns itself with measuring and analysing science.
- **Bibliometrics** is the use of statistical methods to analyze quantitative information pertaining to books, articles and other publications.

Through ‘Open Approaches’ to the scientific record, we can efficiently conduct scientometric reviews.

Studying Inequalities through Open Science

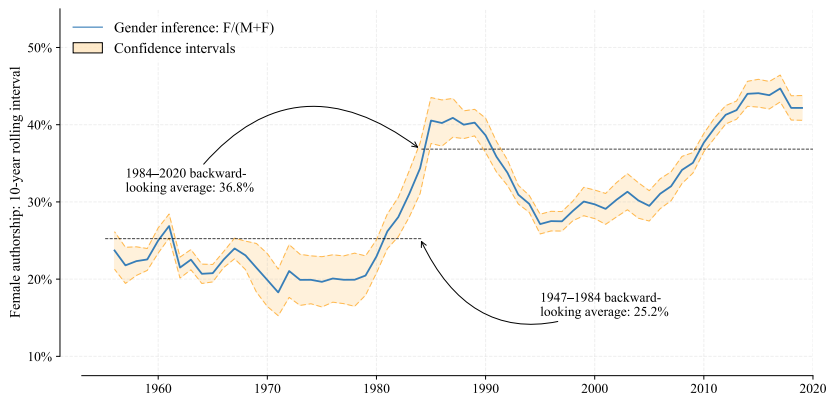
- My early scientometric work analyzed GWAS.
- It took curated lists of catalogued information from the European Bioinformatics Institute. We published two papers which:
 1. Conduct a thorough scientometric review.
 2. Launch a substantial interactive dashboard.
- There is **HUGE** scope for 'Care' research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



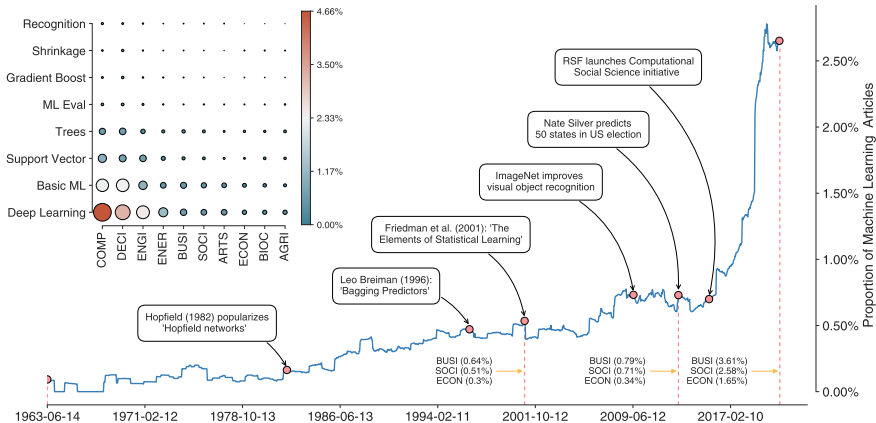


‘The GWAS Diversity Monitor Tracks Diversity in Real Time’
with Melinda C. Mills

Studying Gender with Open Science Tools



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



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

With Dave Kirk and Mark Verhagen

Administrative Data is an invaluable resource for Good.

ADR UK (2018)

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

Administrative Data is an invaluable resource for Good.

ADR UK (2018)

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


[Open Access Home](#)
[Other versions ▾](#)
[DONATE](#)

2,623
out of 74,435
meetings

SEARCH BY LOBBYIST, HOST, SUBJECT

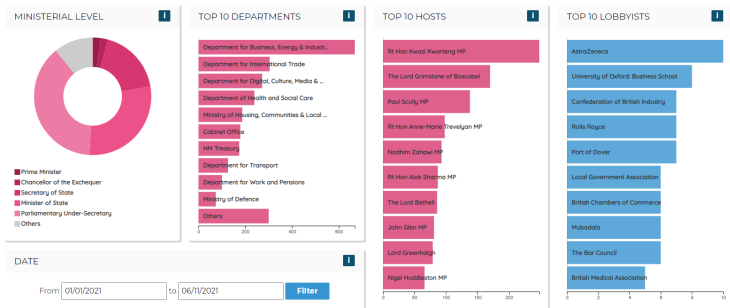
Reset




About ⓘ

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...

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

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

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

Expenditure Over Threshold

38 / 59

Expenditure Over Threshold

38 / 59

A3131

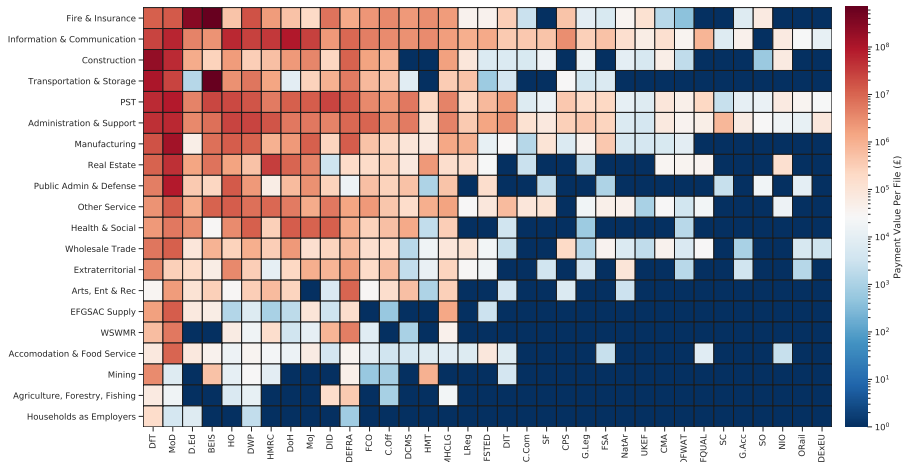
Expenditure Over Threshold

Department	Category	City	Code	Activity	Business Sector	Supplier	Transaction number	AP Amount	VAT registration number	Purchase Invoice number
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Clinic Public Sector	COMMUNITY SERVICES	BEXLEY COUNCIL	11599633	666 000 000	6100071665
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Clinic Public Sector	COMMUNITY SERVICES	BEXLEY COUNCIL	11599635	666 000 000	6100071661
Department of health	NHS	Bexley	CGO	1054261	Other Gen Supplies & Inv	COMMUNITY SERVICES	BEXLEY COUNCIL	11599637	666 000 000	6100071669
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Clinic Public Sector	WYLLIE MASON	BEXLEY COUNCIL	11599642	64 000 000	6100071687
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	UNIVERSITY COLLEGE LONDON HOSPITALS NHS FOUNDATION TRUST	11607291	68 226 000	605485667
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING - NON ACUTE	NHS SOUTH EAST SUSSEX PROJECT	11608000	66 226 000	6100071688
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Voluntary Sector	COMMUNITY SERVICES	BEXLEY COUNCIL	11615604	30 000 000	6100071693
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Independent Sector	ACUTE COMMISSIONING	IN-HEALTH LTD	11615704	38 526 240	787424655
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	NEWWAY NHS FOUNDATION TRUST	11615743	254 250 000	HL00556682
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Voluntary Sector	HOSPICES	GREENWICH & BEXLEY COMMUNITY HOSPICE	11623337	218 000 000	871
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	LEWISHAM & GREENWICH NHS TRUST	11636201	4 440 144 000	9446257
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PROGRAMME PROJECTS	NHS PROPERTY SERVICES LTD	11641857	21 575 140	18402594
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PROGRAMME PROJECTS	NHS PROPERTY SERVICES LTD	11642864	25 672 53	48413715
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	BARFORD AND GRAVESEND NHS TRUST	11642923	-30 000 000	200004184
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11643000	20 000 000	6100071639
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11654425	-3 626 119 34	CM00218970
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	AMBUANCE SERVICES	LONDON AMBUANCE SERVICE NHS TRUST	11663375	26 332 445	BN10662
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11667293	587 757 29	605481192
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11669893	768 536 000	227 489
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	FUNDED NURSING CARE	BURA CARE HOMES	11669922	56 520 340	233561
Department of health	NHS	Bexley	CGO	1054261	Cont Care Adult 100% Fully Funded	RECHARGES	NHS SOUTH EAST CSU	11669995	85 249 000	GB054442045
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11673003	278 325 000	1206094
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	MOORFIELD EYE HOSPITAL NHS FOUNDATION TRUST	11680115	82 568 000	CO0025490
Department of health	NHS	Bexley	CGO	1054261	Hilton-Foundation Trusts	COMMUNITY SERVICES	OXLEAS NHS FOUNDATION TRUST	11701702	228 969 000	37856
Department of health	NHS	Bexley	CGO	1054261	Hilton-Foundation Trusts	COMMUNITY SERVICES	OXLEAS NHS FOUNDATION TRUST	11701707	141 917 000	37932
Department of health	NHS	Bexley	CGO	1054261	Hilton-Foundation Trusts	COMMUNITY SERVICES	OXLEAS NHS FOUNDATION TRUST	11701710	807 726 000	37992
Department of health	NHS	Bexley	CGO	1054261	Income Mx < 11%	BALANCE SHEET	HMC	11720058	47 733 41	604462331
Department of health	NHS	Bexley	CGO	1054261	Income Mx < 11%-Student Loans	BALANCE SHEET	HMC	11720059	738 000	604462331
Department of health	NHS	Bexley	CGO	1054261	National Insurance - 1 vNE-RMS	BALANCE SHEET	HMC	11720065	29 114 84	604462331
Department of health	NHS	Bexley	CGO	1054261	National Insurance - 1 vNE	BALANCE SHEET	HMC	11720066	19 284 18	604462331
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11724081	213 255 000	2007026
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11724088	588 141 000	229584
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11724090	27 120 000	1000041308
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11722209	204 338 000	231364
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11724114	242 139 000	2007026
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Clinic Public Sector	LEASER RES EASY	LEASER RES EASY	11734000	200 000 000	2007026
Department of health	NHS	Bexley	CGO	1054261	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	Bexley	CGO	1054261	Health Svc Res P8th Trust-Over Under Performance	ACUTE COMMISSIONING	MOORFIELD EYE HOSPITAL NHS FOUNDATION TRUST	11738269	30 540 000	CO0025498
Department of health	NHS	Bexley	CGO	1054261	Hilton-Foundation Trusts	COMMUNITY SERVICES	OXLEAS NHS FOUNDATION TRUST	11748440	184 965 000	17541
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	QUEEN VICTORIA HOSPITAL NHS FOUNDATION TRUST	11748386	47 420 000	201009982
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	ARTS HEALTH TRUST	11771725	26 122 000	261232
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	BARFORD AND GRAVESEND NHS TRUST	11771733	2 691 562 000	1000041308
Department of health	NHS	Bexley	CGO	1054261	Hilton-Foundation Trusts	MENTAL HEALTH CONTRACTS	SOUTH LONDON AND MAUDSLAY NHS FOUNDATION TRUST	11777627	25 202 000	80007387
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Independent Sector	COMMUNITY SERVICES - NON ACUTE	BEXLEY COUNCIL	11777630	66 000 000	6100071690
Department of health	NHS	Bexley	CGO	1054261	Hilton-Foundation Trusts	MENTAL HEALTH CONTRACTS	OXLEAS NHS FOUNDATION TRUST	11782402	1 728 324 000	37897
Department of health	NHS	Bexley	CGO	1054261	Hilton-Foundation Trusts	MENTAL HEALTH CONTRACTS	OXLEAS NHS FOUNDATION TRUST	11801176	36 000 000	37806
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	PROPERTY SERVICES LTD	11801413	51 209 19	484138351
Department of health	NHS	Bexley	CGO	1054261	Hilton-Foundation Trusts	CHILD AND ADOLESCENT MENTAL HEALTH	OXLEAS NHS FOUNDATION TRUST	11818020	258 781 000	37805
Department of health	NHS	Bexley	CGO	1054261	Hilton-Foundation Trusts	CHILD AND ADOLESCENT MENTAL HEALTH	OXLEAS NHS FOUNDATION TRUST	11818450	55 506 000	37804
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Clinic Public Sector	HEALTHY SUPPORT SERVICES	OXLEAS NHS FOUNDATION TRUST	11818100	118 000 000	9000019263
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Voluntary Sector	DAY SERVICES	GREENWICH & BEXLEY COMMUNITY HOSPICE	11818190	276 741 000	589
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Clinic Public Sector	DAY SERVICES	DAY SERVICES INTERNATIONAL	11818844	25 000 000	605426016
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Clinic Public Sector	ACUTE COMMISSIONING	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11824000	1 960 531 000	243431
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Clinic Public Sector	LEARNING DIFFICULTIES	BEXLEY COUNCIL	11854467	250 000 000	6100071247
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	ARTS HEALTH TRUST	11856156	1 197 126	261232
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Independent Sector	COMMUNITY SERVICES	BELLEVUE ROAD SURGERY	11963391	66 592 39	BA-032015
Department of health	NHS	Bexley	CGO	1054261	Healthcare Svc Res CCG	NHS 111	NHS BROMLEY CCG	11963393	71 147 88	1002101149
Department of health	NHS	Bexley	CGO	1054261	Clinical/Medical-Independent Sector	COMMUNITY SERVICES	CHAPMAN TOWN SURGERY	11963394	26 703 000	6100071691
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11963373	1 137 848 98	234487
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	GUY'S & ST THOMAS HOSPITAL NHS FOUNDATION TRUST	11963140	1 455 963 000	604923417
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	GUY'S & ST THOMAS HOSPITAL NHS FOUNDATION TRUST	11963387	3 133 848 98	604923417
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Non Contract	PLANNED CARE	LEWISHAM & GREENWICH NHS TRUST	11963381	-129 548 000	8466005
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11963120	1 187 569 000	232736
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11963121	200 156 000	CM00218929
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11963107	229 254 29	232340
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	PLANNED CARE	KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST	11963100	-411 000 000	CM00218919
Department of health	NHS	Bexley	CGO	1054261	Cont Care Adult 100% Fully Funded	CHC AD FULL FUND PERKS HLTH FUND	BEXLEY COUNCIL	11963374	269 574 97	6100071686
Department of health	NHS	Bexley	CGO	1054261	Health Svc Res P8th Trust-Contract Baseline	ACUTE COMMISSIONING	UNIVERSITY COLLEGE LONDON HOSPITALS NHS FOUNDATION TRUST	11963329	68 226 000	605485667

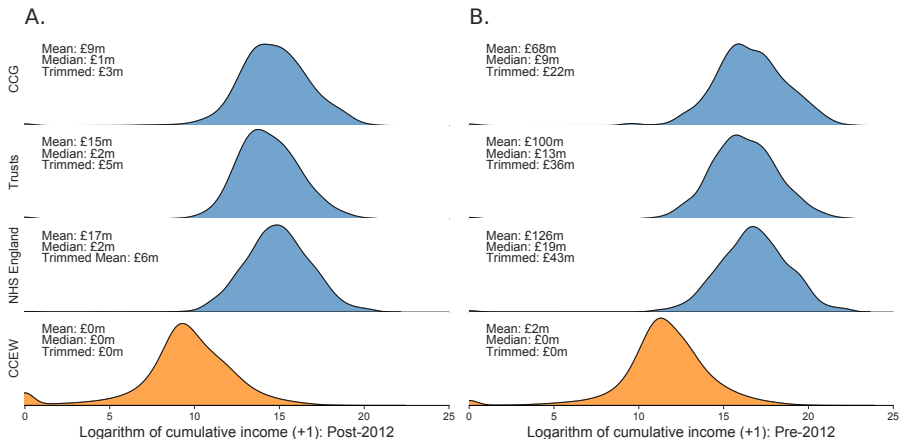
Expenditure Over Threshold

38 / 59

Fun!	Introduction	Scientometrics	Civic Technology	Open Modelling	PRNGs	Conclusion
ooooo	oooooooooooooooooooooooooooo	oooooo	ooooo●oooooo	oooooooo	ooooo	oo

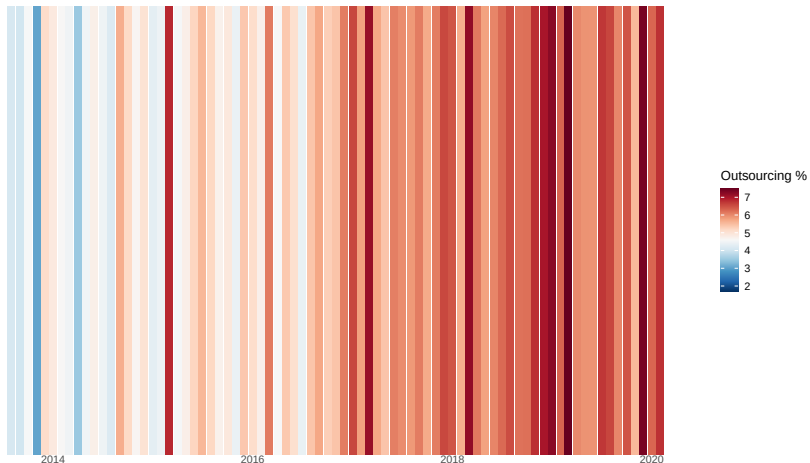


'Tools for Transparency in Central Government Spending'



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

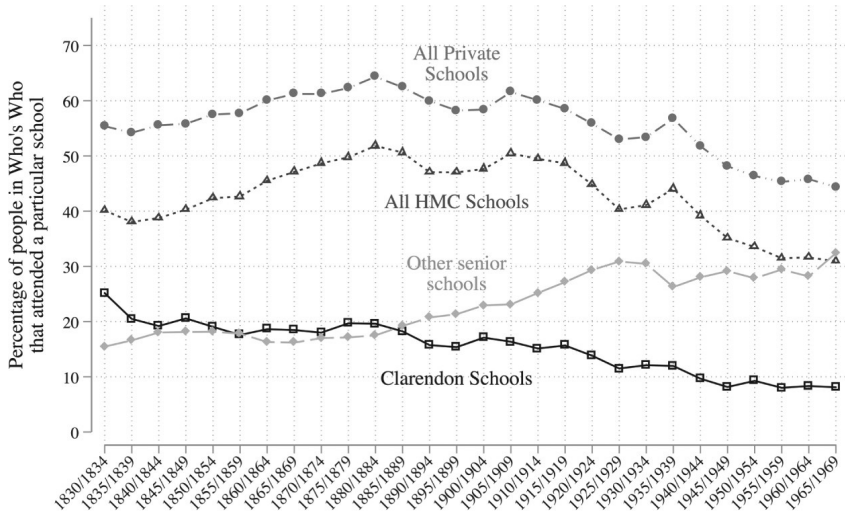
Creeping Privatisation in England's NHS



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

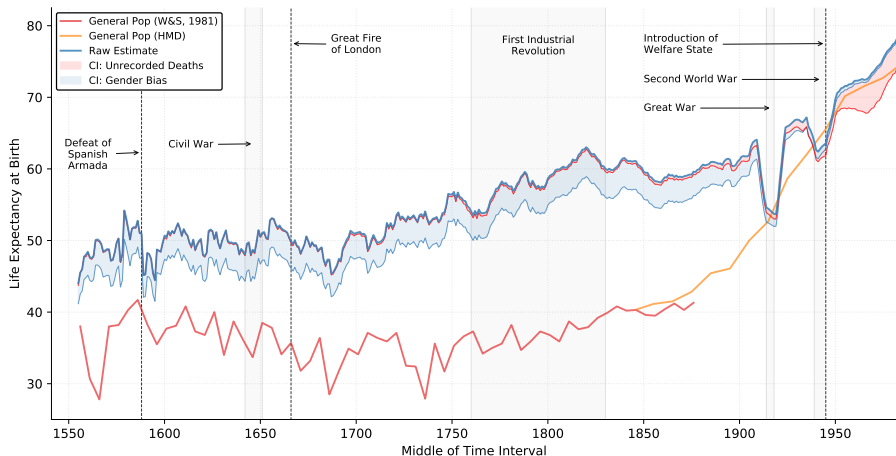
‘Creeping Commercialization in the NHS’

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 Tropsch and Darryl Lundy

68	ABBOTT Henry Bernard of 22 Isley / broad Stroud Green Middlesex died 18 April 1907 at Isley, Longfield Surrey (Engl) Harriette Abbott widow. Effects £16. 6d.	
10	ABBOTT Isaac of Thorpston Northamptonshire bootmaker died 16 October 1908 Administration Peterborough 13 December to Sarah Hannah Abbott widow. Effects £221 19s.	
8	ABBOTT James of The Elms Wood Grove-road Ramsgate licensed victualler died 15 February 1908 Probate Canterbury 17 April to Elizabeth Abbott widow. Effects £284 14s. 2d.	
12	ABBOTT John of 13 Leman-road Leytonstone Essex died 29 January 1908 Administration London 3 March to John Benjamin Abbott contractor. Effects £449 17s. 2d.	
8	ABBOTT John of Betchford Lincolnshire died 15 February 1908 Administration London 25 February to Thomas Abbott farmer. Effects £293 4s.	
12	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 £1916 13s. 4d.	
8	ABBOTT John White of 2 Westbury-gate Hanger Lane Ealing Middlesex died 21 April 1908 Probate London 27 June to Gertrude Agnes Abbott widow. Effects £2139 15s. 2d.	
3	ABBOTT Joseph of 8 Progression-croft, Clapham Field Bawtensall Lancashire organist died 7 November 1908 Administration London 19 December to Maria Jane Abbott widow. Effects £432 14s. 2d.	
10	ABBOTT Lucia Augusta of Banisterrie Lowland Islands (deceased) died 8 June 1908 Administration (with Will) London 12 January to Page Augusta Wade nass clerk attorney of the said Abbott and Anne Margaret Staines Abbott spinners. Effects £1717 3s. 2d.	
10	ABBOTT Mary of Colsterworth Lincolnshire widow died 13 December 1908 Probate London 14 December to Charles Alfred Ernest Abbott bank clerk and George Edward Abbott bank manager. Effects £1942 15s. 4d.	
10	ABBOTT Matilda of Heathcote Court, Arnhem Hill End Devonshire (deceased) died 18 May 1908 Probate London 3 December to Elizabeth Abbott widow. Effects £149 18s. 10d.	
10	ABBOTT Percy Willard of 33 Ashchurch Park Villa Rowan-court Park Middlesex died 5 May 1908 at sea. Administration London 1 August to J. Arthur Abbott chemist's manager. Effects £543 15s. 2d.	
10	ABBOTT Samuel of Cornwell Yorkshire died 21 June 1908 Administration (with Will) York 8 December to George Albert Taylor brewer and Christopher Arthur Baskby millers. Effects £43.	
10	ABBOTT Samuel Mitchell of Elshfold Faversham-road Brevintown Cambridgeshire land and August 1908 Administration (with Will) London 21 December to Elizabeth Abbott widow. Effects £152.	
52	ABBOTT Sarah of Dauntsey Somersetshire spinster died 10 December 1908 Administration (with Will) London 23 April to Emily Gwilder widow. Effects £55 3s. 10d.	
69	ABBOTT Sarah of 11 Springmeadow Bedford Portsmouth widow died 28 April 1908 Probate London 6 May to Eliza Taylor (wife of Albert Taylor). Effects £157.	
10	ABBOTT Sarah of The Square Bedford Northamptonshire (wife of Thomas Abbott) died 17 June 1908 Probate London 29 July to Thomas Abbott engineer. Effects £92 6s. 6d.	

Ongoing research (With Aaron Reeves and Naomi Muggleton) uses and 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 on HPC in parallel: **Done!**
4. Build lifetables to calculate e_0 by wealth at death...: **Pending!**

How is this relevant to Care?

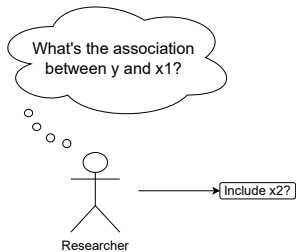
Because quantitative/computational approaches to ‘Care’ research are rare, there is **enormous** potential. For example:

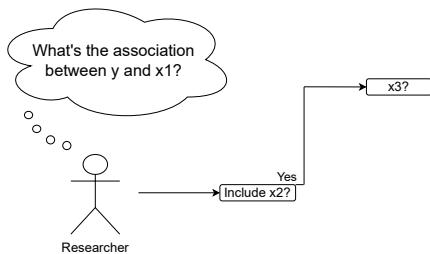
- Analysing provision of hospice/care services through open procurement data
- Scraping CQC data for open source modelling
- Analysing political discourse surrounding care in (open) Hansard data
- Building care-related dashboards via UKHLS or (open) ONS survey data.
- Calculating healthy life expectancies via open source tools!

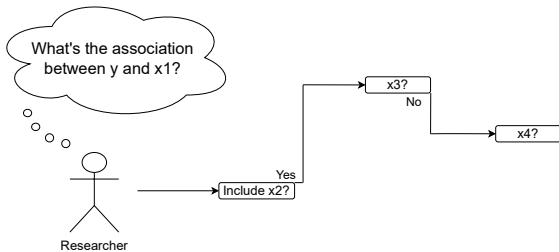
- There are close links between ‘Open Science’ and traditional modeling paradigms!
- By **opening** research, we can examine the certainty of statistical results.
- Does β_1 really equal 7.12452?

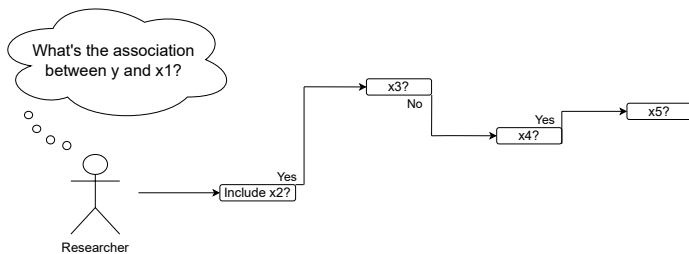
- There are close links between ‘Open Science’ and traditional modeling paradigms!
- By **opening** research, we can examine the certainty of statistical results.
- Does β_1 really equal 7.12452?
- How can we "take the ‘con’ out of econometrics"?

- 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!

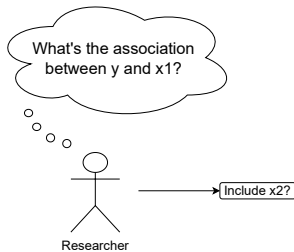


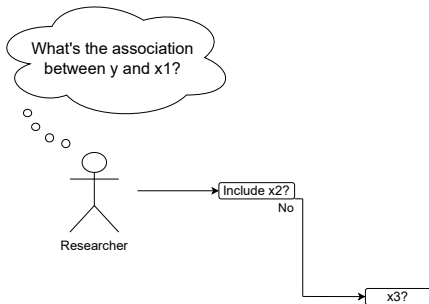


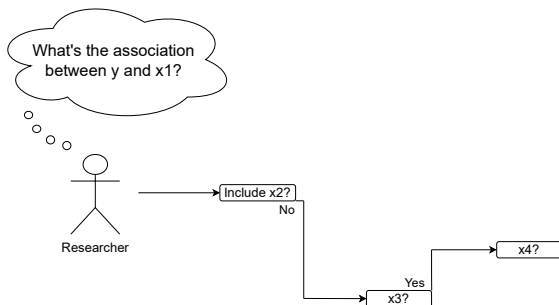


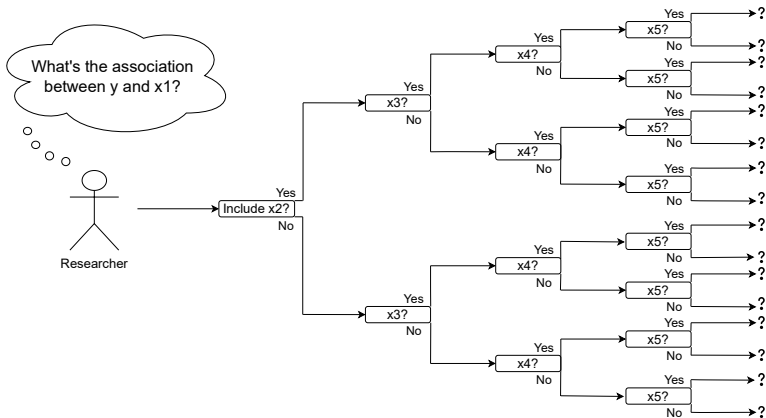




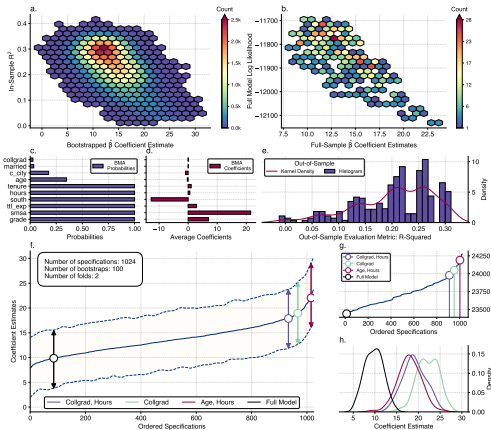








CfC research project with DVI and JY



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:

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

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

¹ESRC Centre for Care

²CIRCLE, University of Sheffield

³Leverhulme Centre for Demographic Science, University of Oxford

⁴Nuffield College, University of Oxford

⁵Oxford Institute of Population Aging, University of Oxford

⁶University College, University of Oxford

⁷Department of Economics, University of Sheffield

⁸School of Social Policy, University of Birmingham

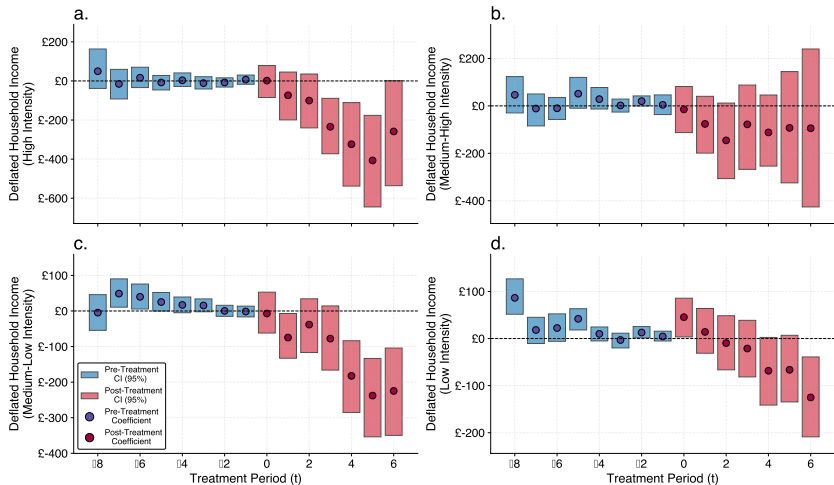
^{*}Denotes joint lead authorship

September 19, 2024

Abstract

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

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




costofcare
Public

 Edit Pins

 Unwatch 2

 Fork 0

 Star 2

main

2 Branches

0 Tags

Add file

Code

MPetrillo25

Create psm

e06ccf1 · 3 days ago

87 Commits

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

README

Cost of Care

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

About

No description, website, or topics provided.

Readme

Activity

Custom properties

2 stars

2 watching

0 forks

Report repository

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Contributors 3



dhvalden Daniel Valdenegro Ibarra

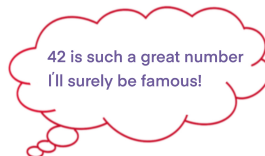


crahal

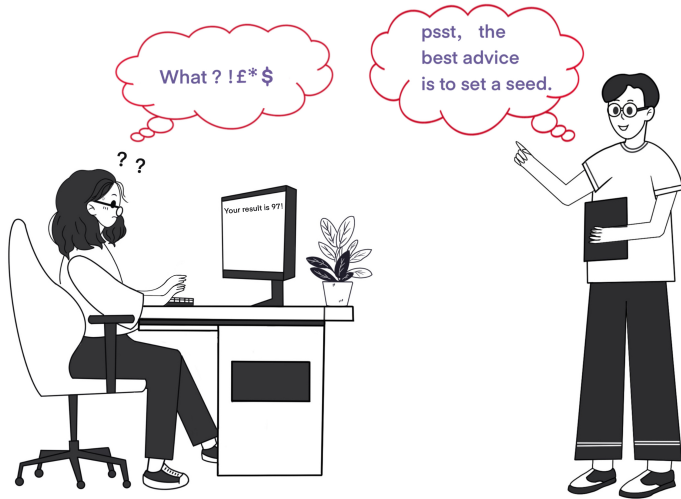


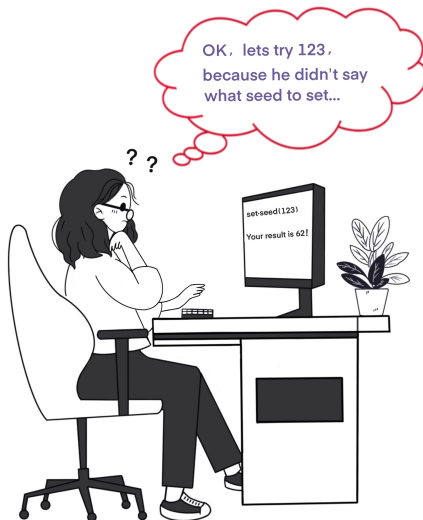
MPetrillo25

















Pseudo-Random Number Generation (Cont.)

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

Pseudo-Random Number Generation (Cont.)

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

Pseudo-Random Number Generation (Cont.)

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

Pseudo-Random Number Generation (Cont.)

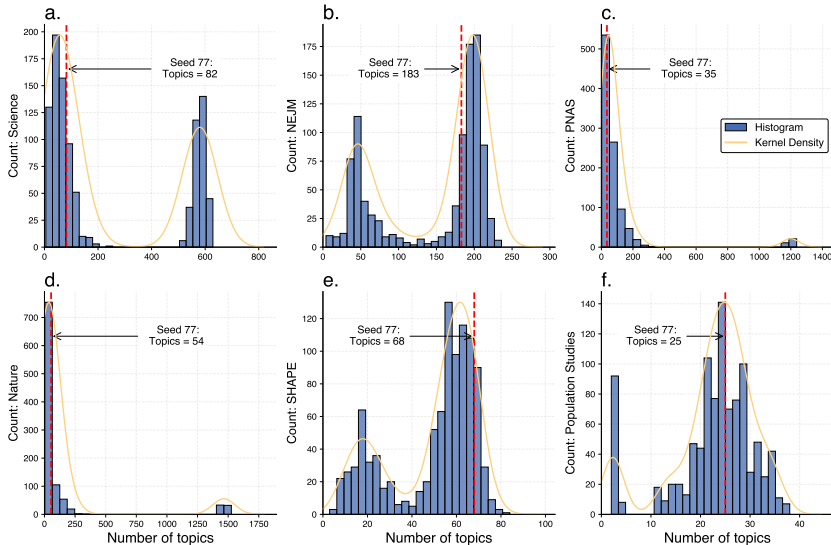
- We want to consider possible variation independent of arbitrary variation in seed choice.
- This is an extremely important and scarcely research problem.
- Pseudo-Random Number Generators occur **everywhere!**
- The variation in estimand can be **huge**.
- We bring attention to this through multiple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

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

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

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





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

Conclusion

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

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

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

Thank you for attending!

Thank you to my collaborators!

Don't hesitate to get in touch at:

charles.rahall@demography.ox.ac.uk

and

github.com/crahall!