Critically Evaluating Statistics examples from Health Inequalities Research

Heather Brown

h.w.brown@lancaster.ac.uk

Why it is important

- In order to be able to make informed decision choices and use evidence it is essential to be able to understand the statistics behind the headlines, health policy, and academic articles

Obese children likely to die up to 20 years earlier than healthy peers

Growing social and economic inequalities across north-west England are directly impacting health

Blackpool is a "hotspot" for inequality and mortality says Nobel Prize winner

Steps to making an informed decision

- Observe your feelings
  - You will have certain feelings when reading public health headlines (e.g. righteous anger, defensiveness, relief)
  - Our reaction will influence how much we buy into the claim and use it in our practice, share with others, or ignore.
  - ‘You must not fool yourself and you are the easiest person to fool’-Richard Feynman

Which of these two headlines would grab your attention?

A) £12 million spent by NHS on diabetics who smoke

B) NHS saves £12 million annually by bankers with private health insurance

Be curious

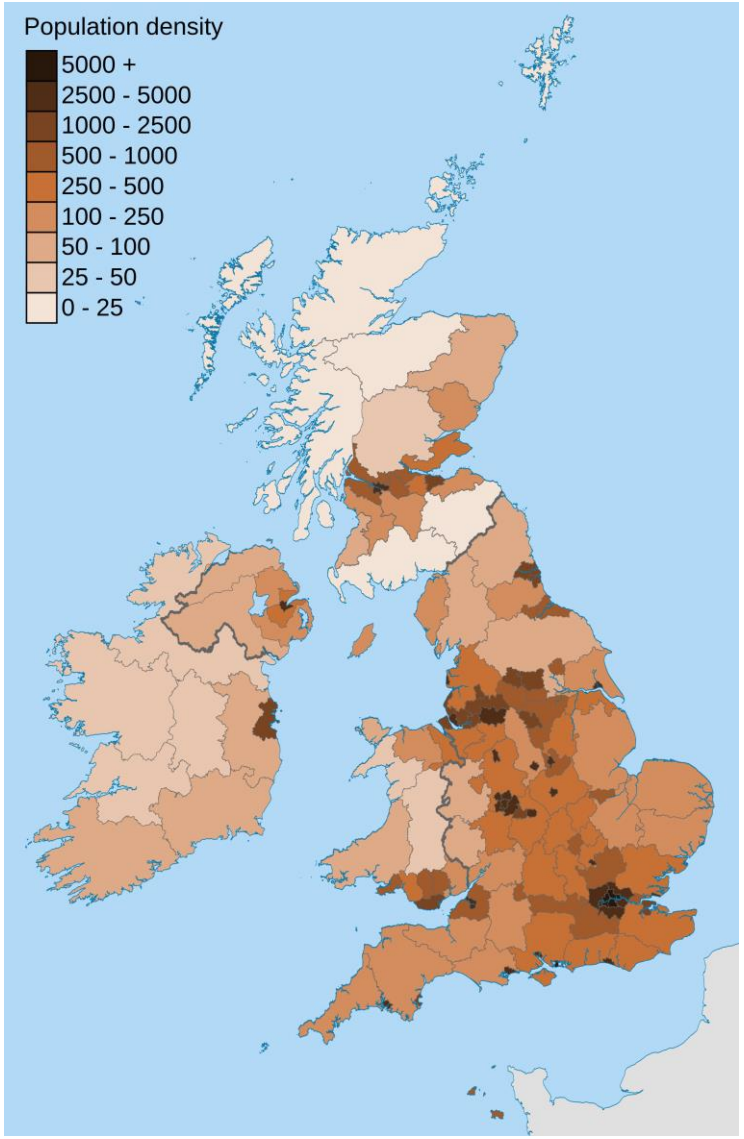
- Try to understand what you are being told
- All statistical statements should make you think
  - Who Claims this?
  - Why?
  - What does this number mean?
  - What is missing

Get the back story

- Of all the statistical claims in the world, you have found this one. Why is that? Where did it come from? Why are you seeing it?
  - Good PR
  - Publication Bias-a study confirming what we already know e.g. smoking causes cancer is unlikely to be published. But, one that shows surprisingly results such as eating chocolate every day reduces the likelihood of developing diabetes is more likely to get media attention.
  - Follow-up studies finding no effect tend not to be published or get no media coverage
  - If a result is surprising or counterintuitive enough it is likely to be wrong-Groucho Marx principle

Putting the results in perspective

- UK throws away 2.5 billion coffee cups a year
- 365 days in a year
- 65 million people in the UK
- This equates to 1 coffee cup per day per person
- Does this seem reasonable?
- How much of total waste is comprised of disposable coffee cups?



Precision of results

- ‘It is better to be vaguely right than exactly wrong’ Read (1898)
  - E.g. problem with election polls in 2015 (margins of error)
- Exaggerated precision has its downside. It can be cumbersome to remember and interpret the data
  - Much easier to remember that the budget of the NHS is approximately £10bn a month
  - Then you can understand if a £50million spending boost will make a difference!

Questions to ask yourself

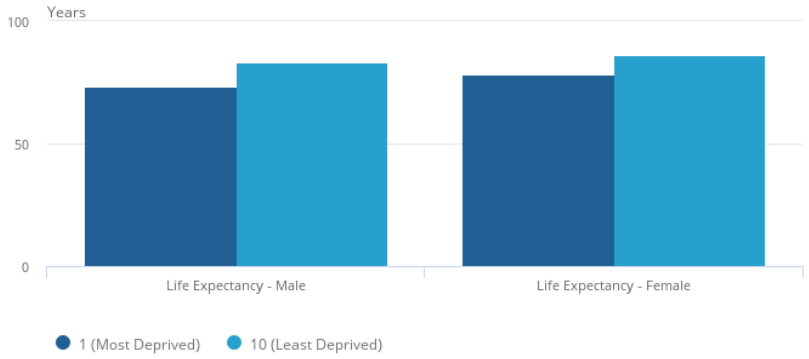
- Understand the claim: *Health Inequality: England's Life Expectancy Gap Is Growing (https://www.forbes.com/sites/katherinehignett/2022/04/29/health-inequality-englands-life-expectancy-gap-is-growing/?sh=61a90d837f2b)*
  - Health Inequalities are rising. What does this mean? How is health being measured? Who is being compared?

So how is the article measuring health inequalities

- Life expectancy
  - 9.7 year gap in life expectancy between men living in most and least deprived areas (increase of 110 days compared to 2015-2017)
  - 7.9 year gap between women in most and least deprived areas (increase of 6 months)
  - Worth noting that 2015 to 2017 also includes Wales-what does excluding Wales means for our interpretation of the findings

Figure 1: Large differences in life expectancy at birth between the least and most deprived areas of England continue for both males and females

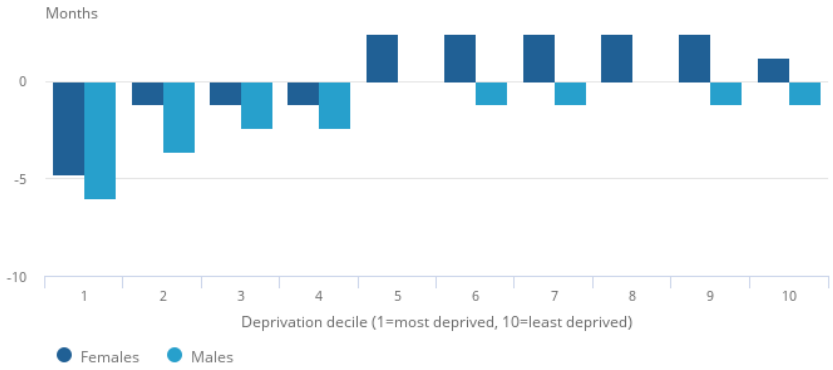
Life expectancy at birth, England, 2018 to 2020



Source: Office for National Statistics

Figure 2: Males and females living in the most deprived areas saw the largest reductions in life expectancy

Change in life expectancy at birth, England, between 2015 to 2017 and 2018 to 2020



Source: Office for National Statistics

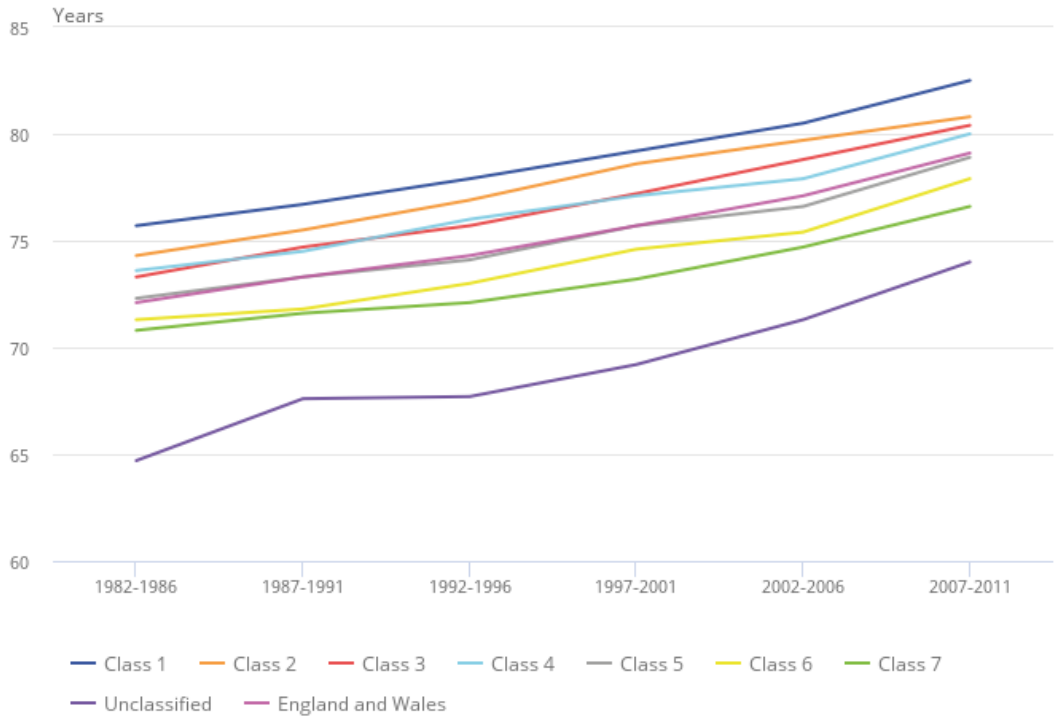
Problem with averages and medians

- Most analysis that you will see will be reporting averages
- Averages can be skewed by a few people at the extremes
- “It is like peering into a room through a keyhole” Sir Andrew Dilnot
- Median tells us about the centre of the distribution but ignores everything else
- If the top and bottom of the health distribution are increasing, but the average sags median health gain will be rising but median health will be falling

Historical Trends: How does the number compare with previous trends?

Figure 1: Male life expectancy at birth for expanded NS-SEC classes including the unclassified and England and Wales, 1982-1986 to 2007-2011

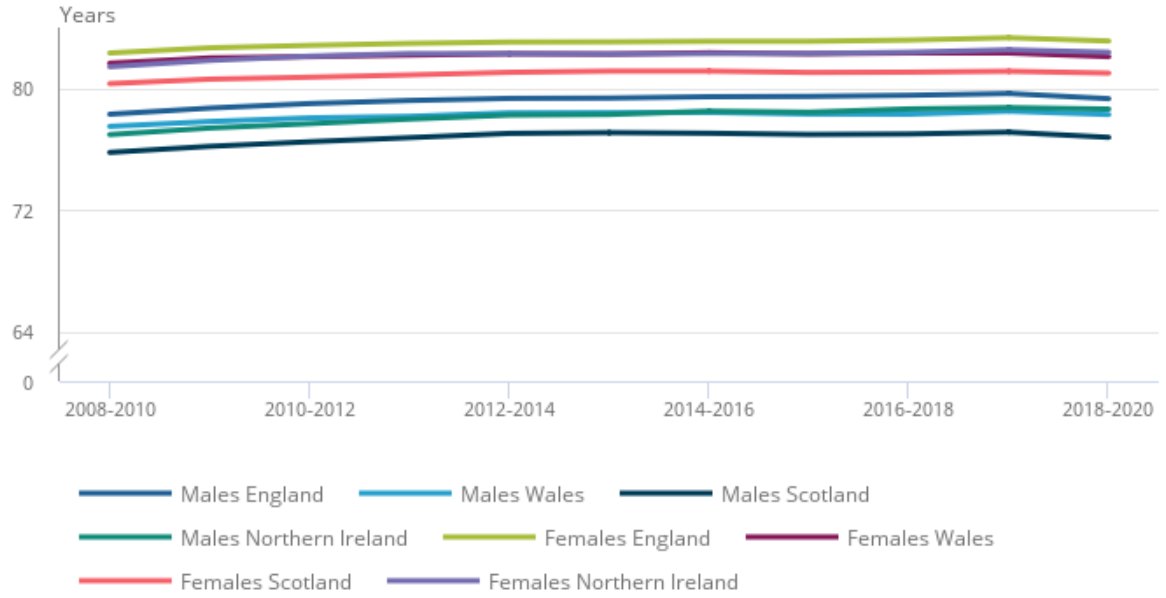
England and Wales, Years



Source: ONS Longitudinal Study

Figure 3: Life expectancy at birth in the UK constituent countries has been increasing more slowly since 2011

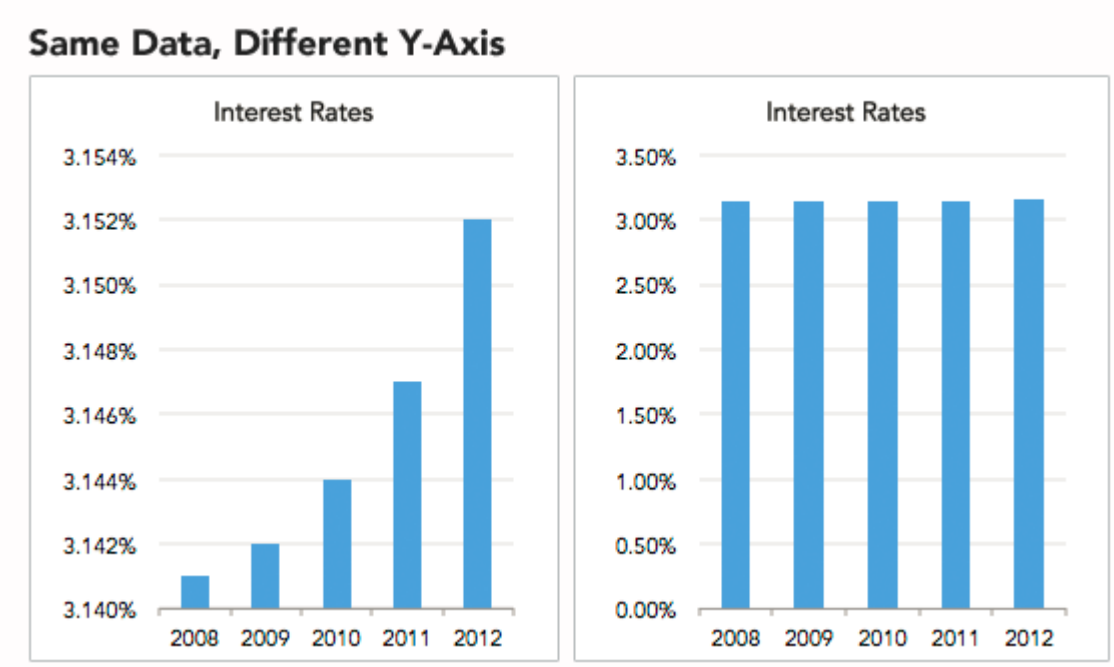
Life expectancy at birth, males and females, UK countries, between 2008 to 2010 and 2018 to 2020



Source: Office for National Statistics – National life tables – life expectancy in the UK: 2018 to 2020

Presentation of Results

- What is being left out
  - E.g. truncating axises to make small changes look big



Beware statistical significance

- A variable can be statistically significant but have no practical importance
- Big data which has large sample sizes and subsequently can take advantage of greater degrees of freedom can mean it is easier for coefficients to pass the hurdle of statistical significance.
- This can be an issue with using administrative or big data

Table IX: Temperature and Outcomes Controlling for Parents' Characteristics in the Census and in the DHS / MIS / AIS

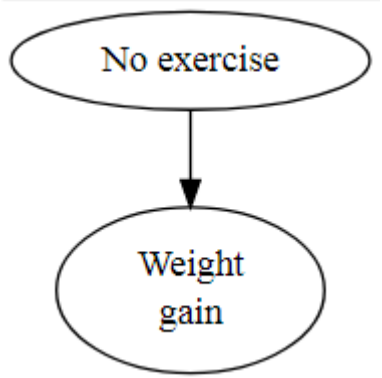
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A. Census								
	Years of Schooling (Attainment)	Years of Schooling (Attainment)	Years of Schooling (Direct)	Years of Schooling (Direct)	Literate	Literate	No Disability	No Disability
Temperature at Birth - 9	0.0228*	0.0262**	0.0418***	0.0439***	0.0115***	0.0118***	6.13e-05	6.22e-05
	(0.0122)	(0.0120)	(0.0117)	(0.0115)	(0.0015)	(0.0015)	(0.0002)	(0.0002)
Mother's Education		0.130***		0.123***		0.0099***		0.0002***
(Years)		(0.0015)		(0.0016)		(0.0002)		(5.73e-05)
Father's Education		0.158***		0.143***		0.0149***		6.61e-05
(Years)		(0.0020)		(0.0018)		(0.0003)		(4.91e-05)
Female		-0.113***		-0.106***		-0.0168***		0.0025***
		(0.0161)		(0.0159)		(0.0018)		(0.0003)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region-Month of Birth FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	913,878	913,878	852,533	852,533	848,743	848,743	898,902	898,902
R-squared	0.340	0.414	0.356	0.425	0.162	0.201	0.010	0.010
Panel B. DHS / MIS / AIS								
	Death	Death	Death	Death	Death			
Temperature at Birth - 9	-0.0056***	-0.0056***	-0.0056***	-0.0053***	-0.0053***			
	(0.0015)	(0.0015)	(0.0015)	(0.0016)	(0.0016)			
Mother's Primary Edu +		-0.0094***						
		(0.0019)						
Mother's Secondary Edu +			-0.0108***					
			(0.0021)					
Wealth Index					-0.0013**			
					(0.0007)			
Year of Birth FE	Yes	Yes	Yes	Yes	Yes			
Region-Month of Birth FE	Yes	Yes	Yes	Yes	Yes			
Observations	86,915	86,915	86,915	90,130	90,130			
R-squared	0.0488	0.0491	0.0491	0.0491	0.0492			

Wilde, J., Apouey, B., & Jung, T. (2014). Heat waves at conception and later life outcomes. *University of South Florida Working Paper*.

Causation vs correlation

Causation

- Is a causal claim being made?
- It is justified?
- Statistics are a summary of a more complicated truth

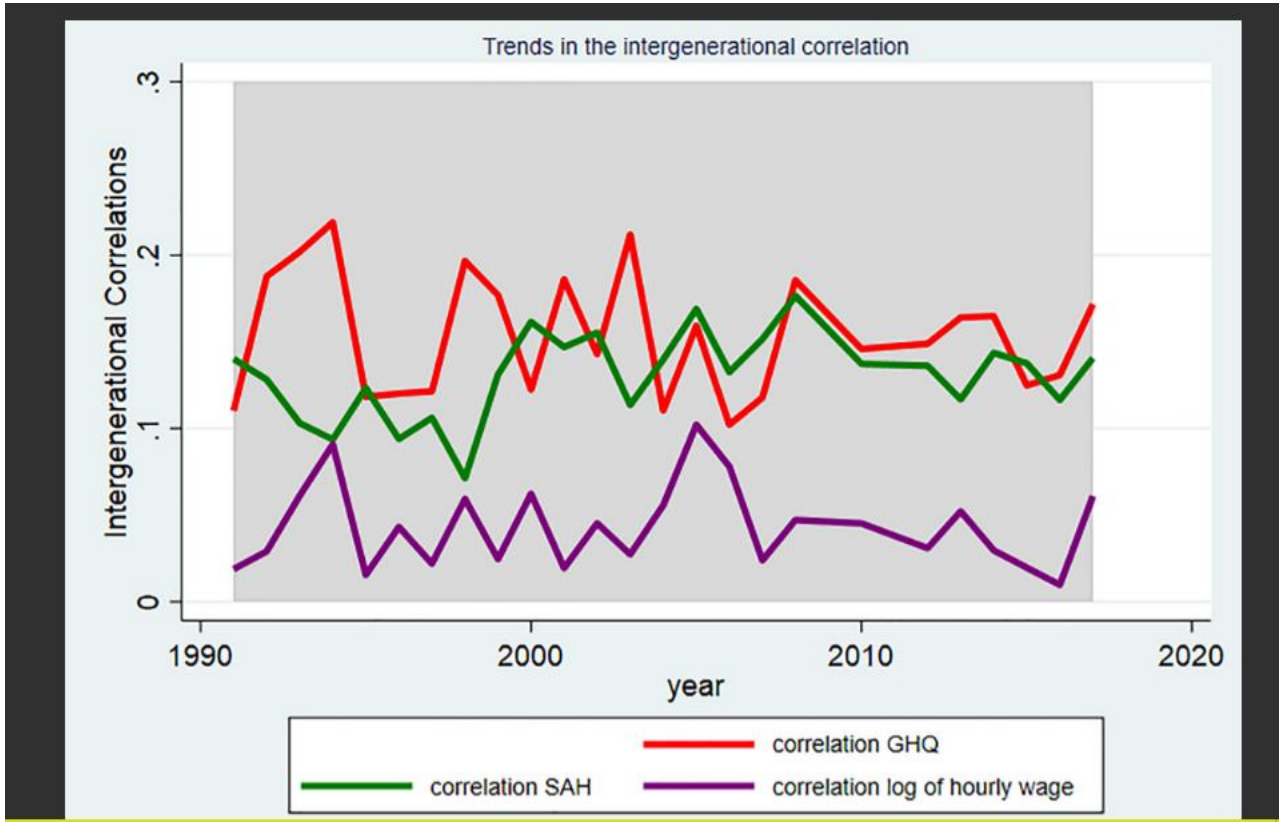


Potential Biases to stop you estimating a causal model

- Omitted Variable Bias: Unobserved factors are correlated with the explanatory variables.
- Measurement Error: There is error in how respondents reply to survey questions. Only a problem if systematic.
- Reverse Causality/Simultaneity Bias: It is not clear if *x* causes *y* or *y* causes *x*.

Putting what we learned into practice

- From: Understanding the role of policy on inequalities in the intergenerational correlation in health and wages: Evidence from the UK from 1991–2017
- What does the following graph show about trends in intergenerational correlations over time



How can we interpret these statistics

We employ fixed effects models on data from the British Household Panel Survey (1991–2008) and its successor the Understanding Society Survey (2009–2017).

Time period	SAH	GHQ	Log of Hourly wage
1991–1998	0.001*** (0.0001)	0.0003 (0.0002)	0.0002 (0.0002)
1999–2009	0.002*** (0.0002)	-0.002*** (0.0002)	0.001**0.0002)
2010–2017	-0.004*** (0.0001)	0.002** (0.0002)	-0.003*** (0.0002)

*** indicates p<0.001

**indicates p<0.05

https://doi.org/10.1371/journal.pone.0234737.t002

More Examples

- From: Financial hardship and health in a refugee population in Australia: A longitudinal study

Next, to estimate the relationship between financial hardship and health over time we employ a dynamic logistic model employing the Mundlak method Mundlak (1978) to proxy for fixed effects in STATA v.15 StataCorp (2017).

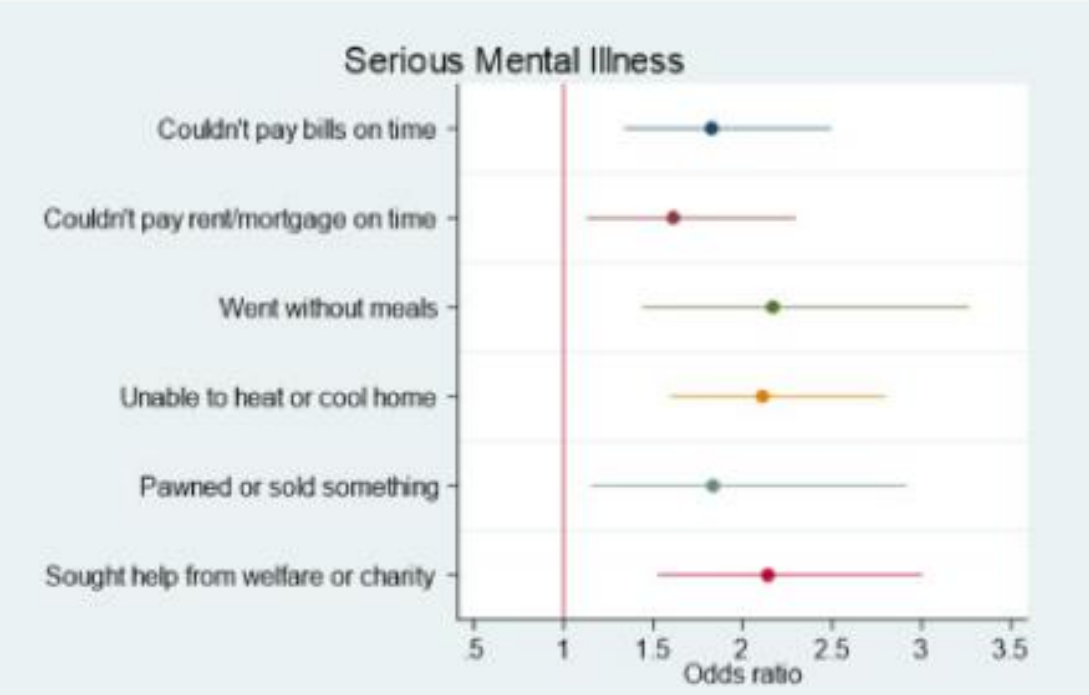
Formally the full model we estimate is:

$$H_{it}^{*} = \beta_1 H_{it-1} + \beta_2 F_{it} + \beta_3 \overline{F}_i + \beta_4 D_i + \beta_5 S_{it} + \beta_5 \overline{S}_i + \beta_6 E_{it} + \beta_7 \overline{E}_i + \varepsilon_{it} \quad (1)$$

$$H_{it} = \begin{cases} 1, & H_{it}^{*} > 0 \\ 0 & H_{it}^{*} \leq 0 \end{cases} \quad (2)$$

Some Results

Mental Illness	Odds Ratio	p-value	[95% Confidence Interval]
Previous Mental Illness	4.88	<0.001	3.57 to 6.65
Age	0.83	0.212	0.62 to 1.11
Female	1.59	<0.001	1.23 to 2.05
Region of Birth			
South-East Asia	Reference		
North Africa and the Middle East	8.93	<0.001	3.42 to 23.32
Southern and Central Asia	4.83	0.001	1.86 to 12.52
Sub-Saharan Africa	6.04	0.003	1.88 to 19.47
Other	3.19	0.317	0.33 to 30.95



References of papers if you want to have a closer look

- Brown, H. (2020). Understanding the role of policy on inequalities in the intergenerational correlation in health and wages: Evidence from the UK from 1991–2017. *PloS one*, *15*(6), e0234737.
- Torlinska, J., Albani, V., & Brown, H. (2020). Financial hardship and health in a refugee population in Australia: A longitudinal study. *Journal of Migration and Health*, *1*, 100030.