Statistics and Intersectionality: how to uncover it in its different forms

Andrew Bell

Sheffield Methods Institute

University of Sheffield

@andrewjdbell

Andrew.j.d.bell@Sheffield.ac.uk

What do we mean by intersectionality

- Different research questions that we could ask
  - Are there generally differences in health across intersections?
    - Eg: are groups different from one another in rates of diabetes?
  - Is a particular intersection different from other intersections?
    - Eg: do Asian, elderly, low-SEP men have particularly high rates of diabetes?
  - Does a particular variable (eg ethnicity) have intersectional effects
    - Is the effect of sex on diabetes rates variable across ethnicities?

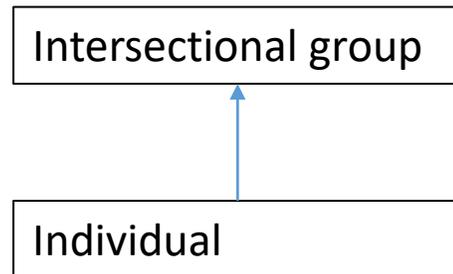# Additive vs Multiplicative effects

- Imagine these levels of advantage

	White	Non-white
Male	10	6
Female	6	2

- Female non-white are most disadvantaged
- But no *additional* effect of being female and non-white
- This doesn't mean that disadvantage isn't real
- But it is additive and not multiplicative

What do we mean by intersectionality

- How do we find these different types of intersectionality?
- One method – multilevel model

Intersectional group

↑

Individual

- Partition the variance… find how much variance at the intersectional level (that is, how big are differences between intersections)
- Then control for main effects – what's left is multiplicative.

But that's not much use for policy makers

- Tells us there is intersectionality, but not *for who.*
- That's a bit more difficult
  - If we make overly-specific intersections, we end up with small numbers in each intersection
    - This also makes policy interventions less efficient
    - And a risk of losing sight of underlying issues of sexism, agism etc
  - Dangers of multiple testing.

- This is not just a quants problem – qual research at risk too!

P values

- As humans we are bad at telling if a pattern we see is actually real (we're possibly evolutionarily designed as such!)

- Risk for qualitative researchers: following the patterns we expect to see

- Quantitative researchers: often use statistical significance / P-values
  - "the probability we would have seen a pattern that clear (or clearer) if actually there was no pattern in the population"
  - Arbitrarily use the threshold of 0.05… but that means 1 / 20 will find such a pattern even when none exists
  - And with intersectionality we're testing much more than 20 intersections

P values

- So a big risk we will find non-real differences in the population
- That means potentially wasting resources on poorly targeted interventions
  - Groups at need get less resources than they should
  - Groups get potentially harmful interventions

- For me: the solution isn't to just to qual (that has the same problem!) but to build theory (from qual) into statistical models so we can ask specific questions
  - From "where are there intersectional effects" to "is X intersection disadvantaged"

Partial solution?

- Multilevel model – through shrinkage. Brings unreliably measured intersections closer to the mean. Reduces multiple testing problem (Jones et al 2016, Evans et al 2018)

- But only complete solution if intersections are independent… which they aren't! (Bell et al, 2019)

- Need to control for main effects, and possibly some interactions (iteratively)


HbA1c (mmol/mol)
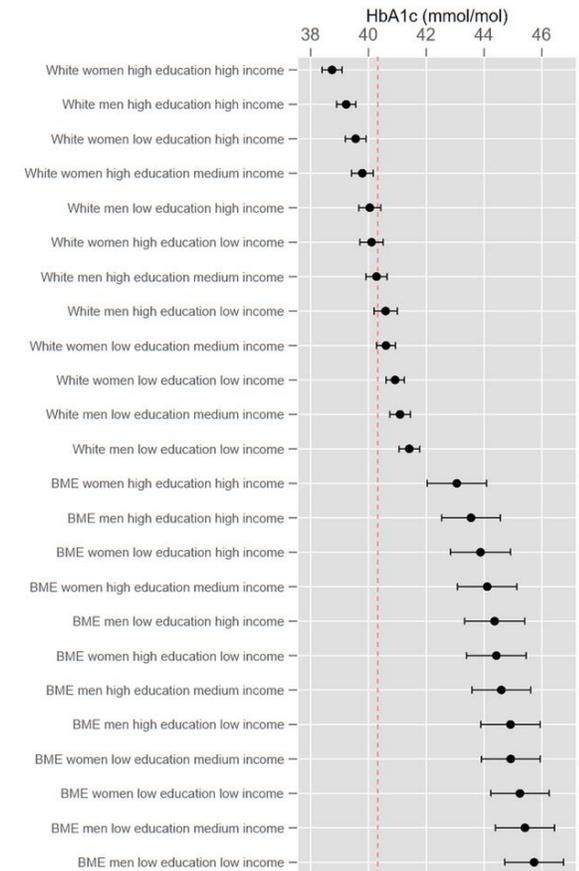
X axis = effect size. Reference level = sample mean.

Other issues – what variables?

- ethnicity, sex, social class, income group, age, gender identity, sexuality, political orientation,

- Different variables require different treatment – continuous (age, income) vs categorical (ethnicity, social class)

- Grouping continuous variables into groups = wasting information

- Within-group variance still dominates usually: the group doesn't define the person

Other issues – what variables?

- ethnicity, sex, social class, income group, age, gender identity, sexuality, political orientation, neighbourhood……
  - End up with very small groups very quickly

- Different variables require different treatment – continuous (age, income) vs categorical (ethnicity, social class)

- Grouping continuous variables into groups = wasting information

- Geography: a separate category?

Intersectional group	Neighbourhood

Individual
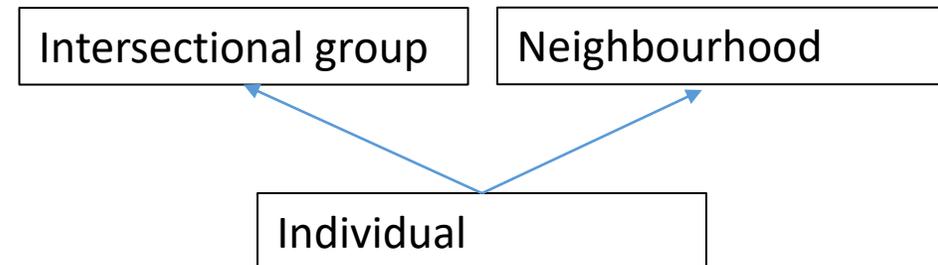
Other issues – what variables?

- My view: needs to be theoretically informed which variables we choose

- What are the processes of discrimination and disadvantage, and who would we expect them to harm?

- Work with qualitative researchers to identify the *processes* so that the model makes sense

Key practical advice

- Base variables on literature / background knowledge

- Use a multilevel model

- Test for main effects, then interactions within the multilevel framework

Summary

- Intersectionality is complex, so the approaches to measure and understand it need to be complex also

- Complexity needs to be grounded in theory

- Requires careful understanding of the research question being answered

- And careful understanding of statistical significance – what it does and doesn't mean

References

- Bell, A., Holman, D., & Jones, K. (2019). Using shrinkage in multilevel models to understand intersectionality: a simulation study and a guide for best practice. *Methodology, 15*(2), 88–96.

- Evans, C. R., Williams, D. R., Onnela, J.-P., & Subramanian, S. V. (2018). A multilevel approach to modeling health inequalities at the intersection of multiple social identities. *Social Science & Medicine, 203*, 64–73.

- Jones, K., Johnston, R., & Manley, D. (2016). Uncovering interactions in multivariate contingency tables: a multi-level modelling exploratory approach. *Methodological Innovations, 9*, 1–17.

Thanks for listening!

Andrew.j.d.bell@Sheffield.ac.uk

@andrewjdbell