

How it's been doing a statistics PhD plus some research

EA Imperial

Adam Howes

Imperial College London

February 2022

Academic background

- Maths undergraduate at Durham, no real ideas about career
 - Worked in a pub first year, Cranfield second year, and Basque Center for Applied Mathematics third year
 - Dissertation about Bayesian optimisation
- Not enough statistics courses at Durham, so did a masters at Warwick
 - Dissertation about Bayesian models for combining sources of evidence
- StatML CDT (Imperial / Oxford) started the year I was applying
 - Somewhat successor to OxWaSP (Oxford / Warwick)

EA background

- Heard about 80,000 Hours in 3rd year from a friend
- Listened to a few podcasts, found them compelling, first time anything career related really resonated with me
- Read a few books over summer, then started going to the EA society at Warwick
 - Events like retreats, more people in the community, helped to add depth
- First EAG I went to was the year I started at Imperial
 - Volunteering can be a good way to get a place

Is it worth doing a PhD?

- Skills or talent gap for many causes in EA
 - Rather than funding gap
- Independent research skills developed during a PhD important for many causes
 - AI, biorisk, global priorities, global health and development
- Can be useful for signalling expertise
 - Then transfer into e.g. policy
- 3/4+ years, possibly large opportunity cost

Any tips about applying for PhDs

- Find (and call if possible) any previous students of the supervisor
 - What are they doing now?
- Assess the vibe among current graduate students
- Be interested in things, read papers, attend reading groups
 - “List of things you don’t need permission to do”
- Doing research as an undergraduate is a good test

Some negative reflections on PhD experience

- I, along with many (all?) of my friends doing PhDs, have had periods where it has negatively effected wellbeing
 - “Grad school is worse for public health than STDs” <https://www.benkuhn.net/grad/>
- Little structure, often sporadic feedback, uncertainty about research direction and progress, narrow contribution (that perhaps few people care about) to a large field, low pay, alignment with supervisor and program, (in some fields) somewhat pyramid scheme-y
- Different to undergraduate: not learning material then passing test for another 3/4 years

What's my PhD about

- HIV/AIDS is an infectious disease which is endemic in sub-Saharan Africa
- Epidemic indicators can be used to guide policy response
 - Prevalence (how many people are infected)
 - Incidence (how many people newly infected)
 - Coverage of treatments (how many people have (don't have) treatment)
- My area of research is Bayesian spatio-temporal statistics, the method we use to perform small-area estimation
 - (I'll explain what these terms mean)

Bayesian statistics

- First step in Bayesian statistics is to write a probabilistic model for the data
 - Focus is on understanding and modelling scientific phenomenon, rather than e.g. remembering arcane statistical tests
- Second step is to perform inference (which is usually done with an algorithm)
 - Given we observed certain data, what can we learn about the parameters of the model?
- Things might go wrong with the above two steps with the data, probabilistic model, algorithm
 - Should also perform checks, following a “Bayesian workflow”

Spatio-temporal statistics

- Say that you have a collection of people who have taken a HIV test, and their results are either positive or negative
- If have the location for these people, this might be important information to include in our probabilistic model
- Similarly, the time that they took the test might also be important
- Spatio-temporal statistics is just when we think these two features of the data are important to take into account
- It allows us to answer spatio-temporal questions
 - What's the situation at time t and location s ?
 - Often important for policy

My work

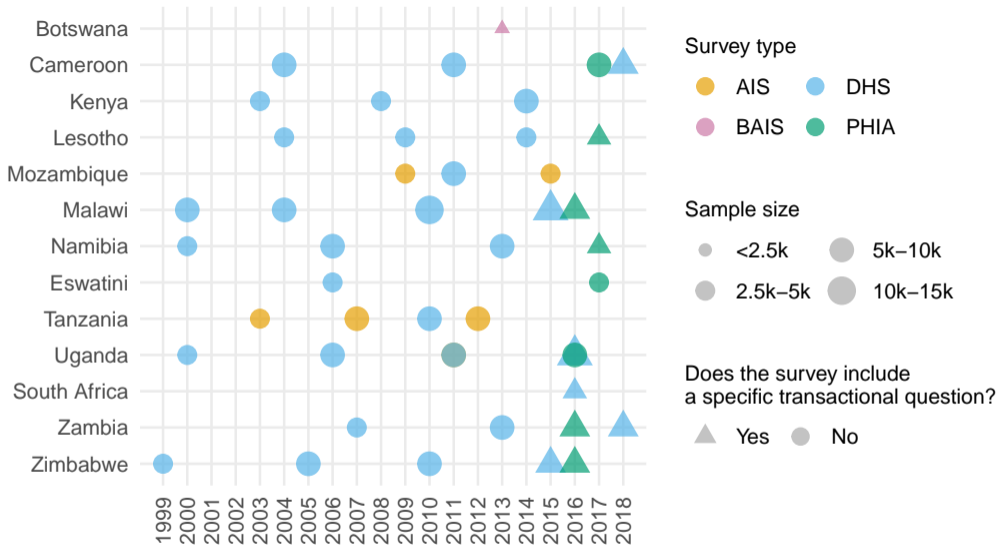
- I'm interested in applied and methodological work
 - Applied: focus is on answering a question about the world
 - Methodological: focus is on developing techniques which can be used to answer questions about the world
- The deliniation between these might not be clear sometimes
 - To answer questions about the world you'll often have to develop new methods
 - To develop new (relevant) methods that you should be thinking about which questions they will allow you to answer to think
- Some examples of my work
 - Applied: estimating numbers of young women at risk
 - Methodological: investigating models for spatial structure

Estimating numbers of young women at risk

- Women 15-29 are disproportionately at risk of HIV
- To efficiently target programs and resources to the most at risk women, we're working on small-area estimates of the number of women in different risk groups

Category	Description	Risk ratio
None	Not sexually active	0
Low	One cohabiting sexual partner	1
High	Non-regular sexual partner(s)	1.72
Very high	Young women from key populations	13

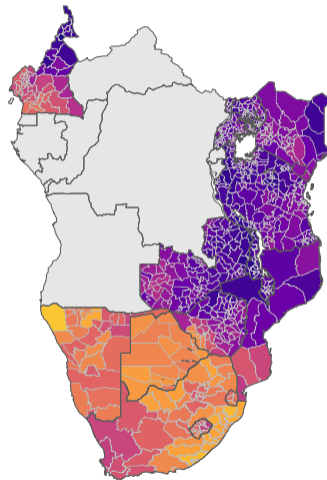
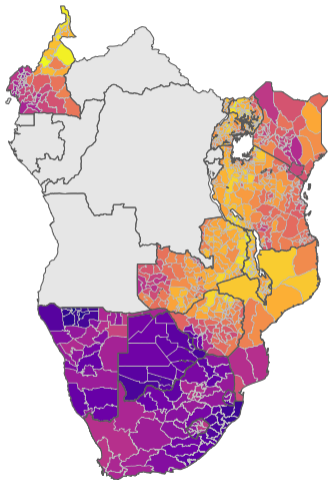
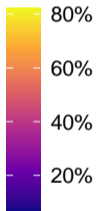
Table 1: HIV risk categories and risk ratios.



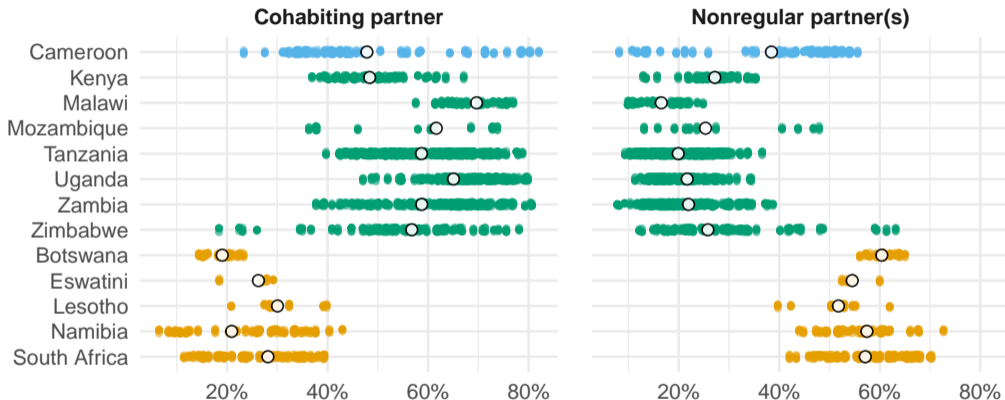
Cohabiting partner

Nonregular partner(s)

Proportion
of women
20–29

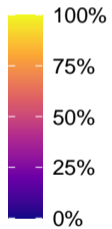


Regions of sub-Saharan Africa ● Central ● Eastern ● Southern

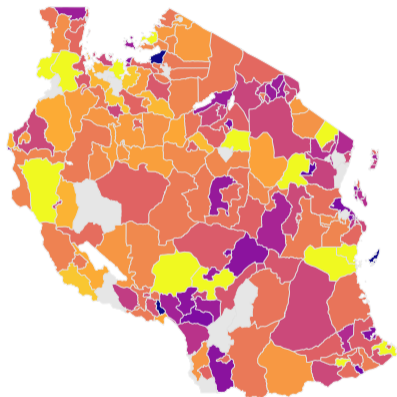


Not sexually active (not shown) + **Cohabiting partner** + **Nonregular partner(s)** + **FSW** (not shown) = 100%

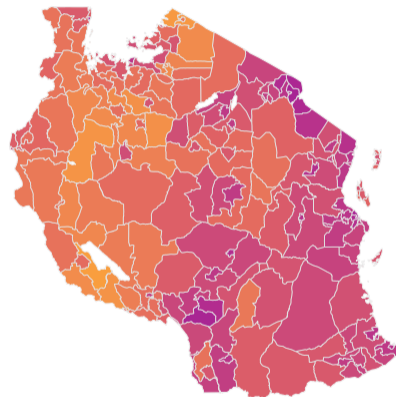
Proportion of women 20–24 cohabiting (2010)



Direct



Modelled



The Besag model

$$\phi_i | \phi_{-i} \sim \mathcal{N} \left(\frac{\sum_{j:j \sim i} \phi_j}{n_{\delta i}}, \frac{1}{n_{\delta i} \tau_\phi} \right)$$

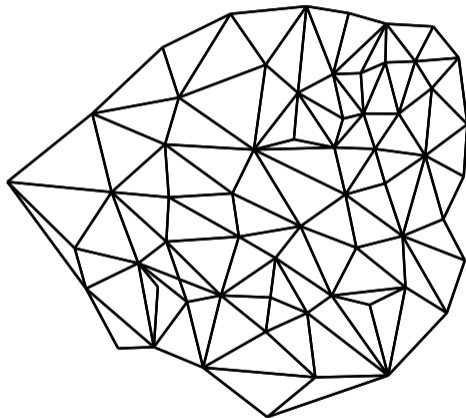
Mean of ϕ_i average of adjacent units and precision is proportional to $n_{\delta i}$

`## Warning in CRS(proj4string): CRS: projargs should not be NULL; set to NA`

Geography



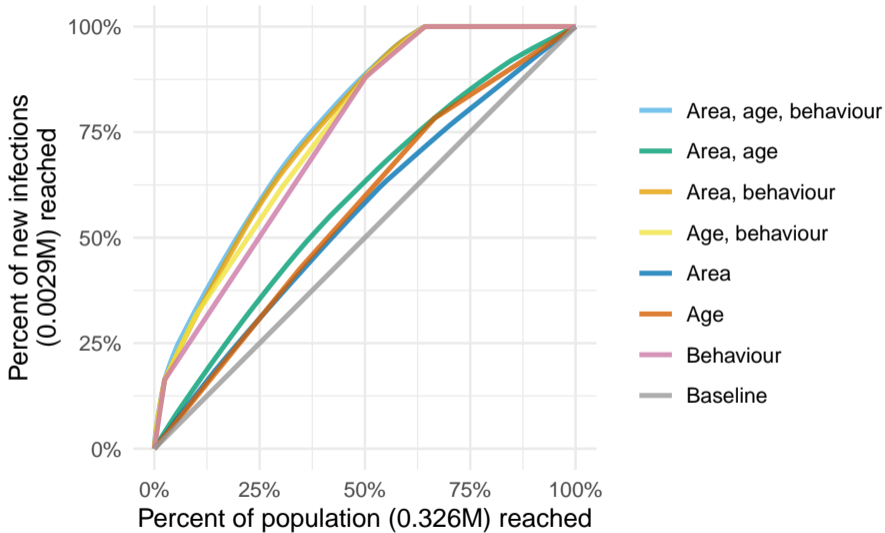
Graph



Impact of this work

- Could be used to guide HIV interventions for women 15-29
- Interventions suggested include
 - STI-screening and treatment
 - Economic empowerment,
 - Comprehensive sexuality education
 - PrEP
 - PEP

Botswana



Some possible routes for the future

- Small-area estimation of health and development indicators
- Outbreak disease surveillance
- Monitoring of the environment (e.g. sea-ice for climate, air-pollution for global health)
- Statistical ecology (as it relates to animal welfare)
- AI safety (likely the more applied FAT-style work than theoretical)
- Predicting (or helping to win) elections
- Meta EA

Current plan

- Outbreak disease surveillance for reducing GCBRs
- Longtermist issues likely score better on neglectedness and scope
- Biorisk cause area “competitive” with AI risk
 - Better personal fit to biorisk
- Biosecurity reading group this term, Fridays at St Mary's 12-1!

Thanks!