Population Sciences

Global Challenges and Opportunities

Sam Clark

Department of Sociology, The Ohio State University, Columbus, USA

Affiliated: MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand

Affiliated: Department of Statistics, University of Washington

Carolina Population Center, UNC

Zoom

February 10, 2022

Three big challenges

- **Climate emergency** – all three affecting and affected by each other:
  - human population – *central role*
  - climate and atmosphere
  - natural environment: land, sea, life
- **Health**
  - aging, continued epidemiological transition
  - epidemics/pandemics – prepare for the next one
  - structural issues affecting health and access to health care – inequalities, population structure
- **Public trust in science**
  - discrediting experts
  - misinformation – wrong
  - disinformation – deliberately deceptive

Role for population scientists in big challenges

- **The big challenges are global – require global response**

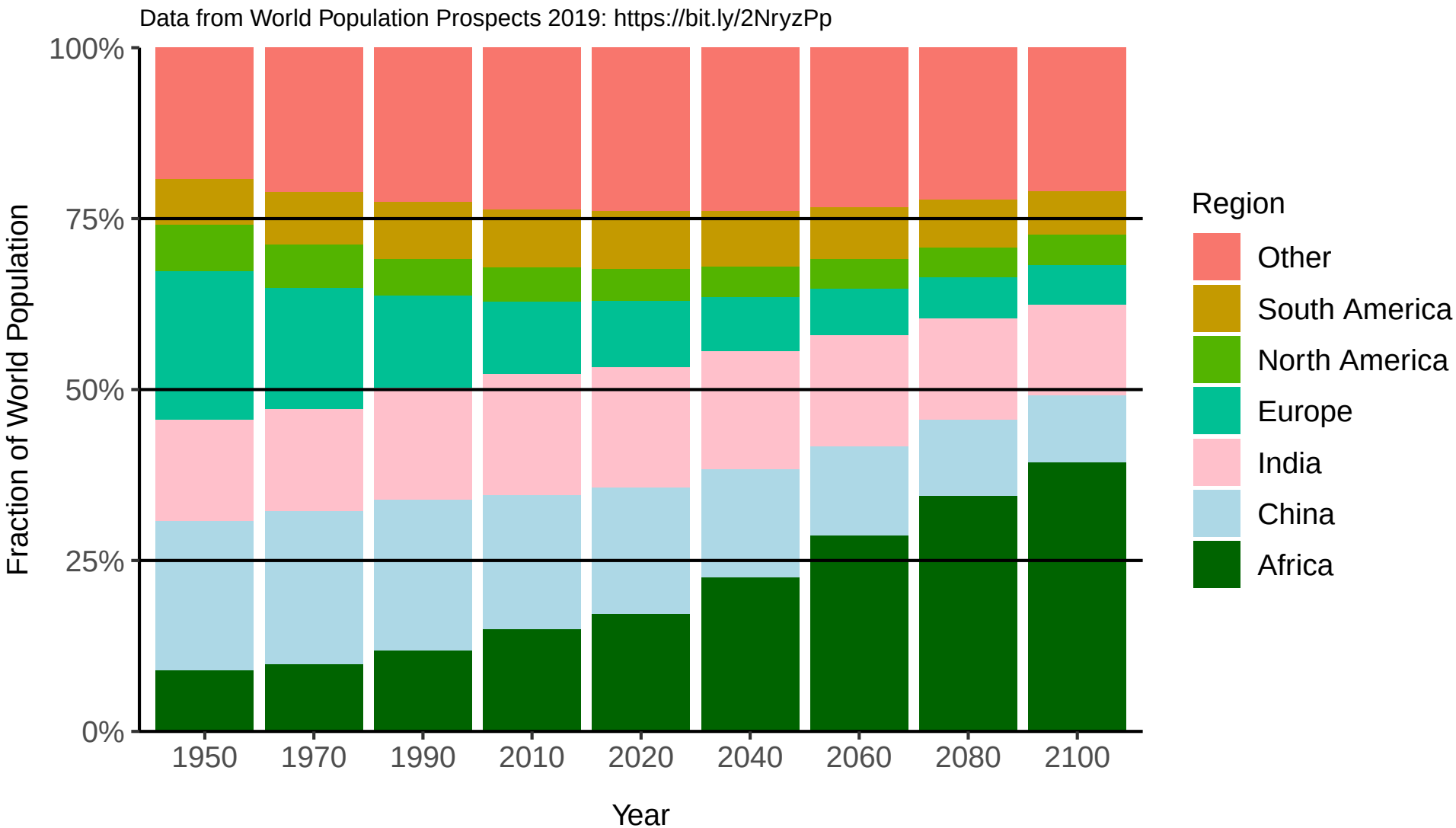
- Understand the role of global and regional population size and dynamics in the big challenges
- Global and regional population forecasting that works with climate, ecology, and economic models $\rightarrow$ planning scenarios
- Careful measurement of population and health processes at global and regional scales
- **Deep integration and collaboration with related disciplines**
  - earth and biological sciences
  - related behavioral sciences
  - health scientists
  - statistics
  - computer science

Vision

1. Develop global approach with a focus on Africa
2. Statistical demography and epidemiology
3. Environment – Population – Economy
4. Communication: population science and the public

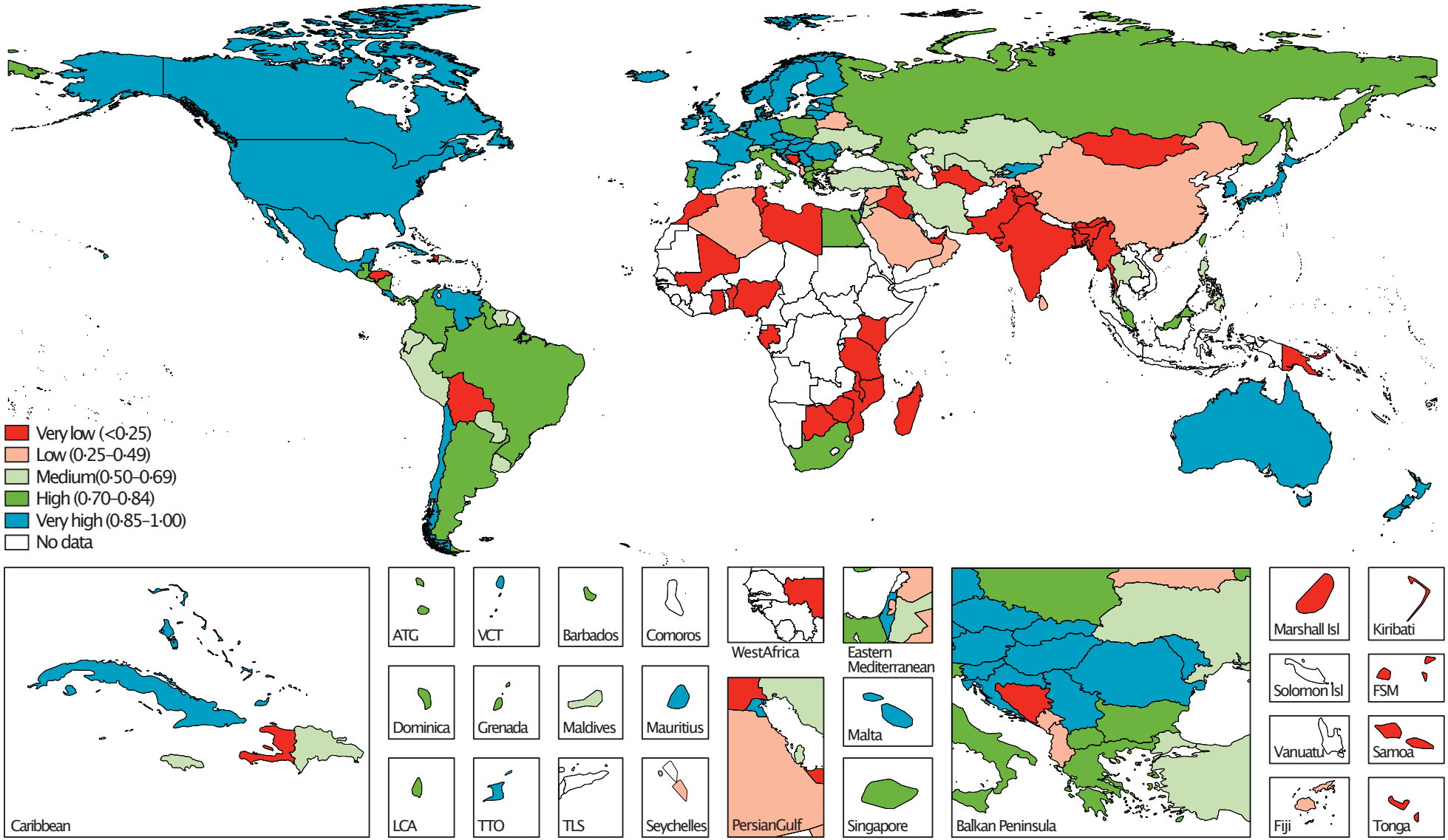
Population composition by region – Africa

(UN DESA Population Division, 2019)



Vital statistics performance index

(Mikkelsen et al., 2015)



Global approach 1

- **The next century is largely about *Africa***
  - Share of the globe
  - CO_2 production
  - Consumption
  - Health: productivity/consumption
- Basic data are missing – gigantic gap in empirical understanding of population and health in Africa and other lower- and middle-income countries
- Reshape North-South relationship – need balance, move center of gravity toward the South
- Develop joint approaches
  - Scientific training and mentoring
  - Research priorities
  - Funding and conduct of research
  - Dissemination and application of research findings

Global approach 2

- More data and fewer models
  - Move data production about the South to the South – see NIH DS-I Africa program for data science in Africa
  - Develop new efficient, integrated data collection/indicator estimation procedures
  - Develop alternative to IHME and the general North-based, North-centric approach
- Global engagement is imperative for health
  - There will be another pandemic – we need to be prepared
  - Cooperative global monitoring
  - Truly global response – cannot continue to have large, untreated reservoirs for pathogens to rapidly evolve new, potentially dangerous variants
- Global aging and chronic disease
  - Most of globe will age rapidly over the coming century
  - Burden of chronic diseases of aging and other aging-related issues will dominate epidemiological landscape (between pandemics!)

Statistical demography and epidemiology 1

Challenge

- Explosion of new data sources: big data, administrative records, digital exhaust, remote sensing, etc.
- Many new data sources are un-anchored, uncalibrated, or unprincipled: e.g. no sampling frame and no easy generalizability
- Traditional demographic methods are deterministic
- Traditional epidemiological methods largely based on sample survey methods
- Global initiatives like MDGs and SDGs require multi-dimensional disaggregation, including space and time
- Complete, disaggregated picture requires many different data sources and extensive modeling, smoothing, interpolation, extrapolation, etc.

Statistical demography and epidemiology 2

Challenge continued

- Realistic quantification of uncertainty, confidence, etc. essential, *including from non-sample-based data*
- New era of computational power and ability to apply Bayesian approaches
- Machine learning and related approaches to automation developing rapidly
- Other fields are very rapidly using advanced methods to explore traditionally demographic and epidemiological questions – e.g. computer science, statistics with CV19
- Advanced methods that rely on other disciplines can make things hard/impossible to comprehend and mask bad science – e.g. IHME
- Need to maintain the replicability, comprehensibility, and hence trust in demographic and epidemiological work

Statistical demography and epidemiology 3

Action

- Must continue to own demographic and epidemiological methods
- Incorporate modern statistics, computer science, and applied math into demographic and epidemiological training
- Encourage a methods-development track with the aim of keeping up with developments in statistics and computer science – Bayesian(-like) methods, new data, new data management, machine learning, etc.
- Continue and strengthen mathematical demography and epidemiology methods development
- Actively seek to hire/support faculty and researchers from statistics, computer science, and applied math to teach and do research in demography and epidemiology
- Emphasize graduate training in these areas to develop new generation of demography and epidemiology researchers who can develop and apply new methods/approaches

Statistical demography and epidemiology 4

Action continued

- Encourage Bayesian approaches
- Include Global South – e.g. NIH DS-I Africa program for data science in Africa
- Develop new data/analytical paradigm: real-time data and analysis – think 'ticker tape' or real-time dashboard of population and health indicators
- Meld data collection and analysis into continuously operating, self-calibrating, self-focusing instrument – think mammalian visual system: scanning, variable granularity, auto-calibrating, sophisticated post-processing, auto-focus, auto-attention shifting
- Be much more ambitious than we currently are – we need big population, health monitoring infrastructure; similar to a Webb Space Telescope or a particle collider

Environment – Population – Economy (EPE) 1

Situation

- Human activity [population] affects the natural world – atmosphere, environment, ...
- The natural world [environment] affects agriculture, living conditions, ...
- The economy (and legal frameworks) mediates / incentivizes / constrains much of human activity
- These three systems work together in a complicated set of effects and feedbacks to generate the climate / environmental emergency we face now
- Economic policy is likely the most direct, effective, and timely way to bring about positive changes
- This is *very* complex
- This is also very urgent and important

Environment – Population – Economy (EPE) 2

Challenge

- Significant bodies of theory and good modeling frameworks exist for each of the three components individually, or in pairs
- There isn't much that links the three, especially good modeling frameworks
- *Urgent need to develop useful models that include realistic links between the three systems* $\rightarrow$
  - Forecast likely futures for the whole system
  - Assess and compare possible interventions
- **Population size and dynamics** are key to the whole system, and therefore, demographers should play a central role

Environment – Population – Economy (EPE) 3

Action

- Build multidisciplinary consortium to design modeling framework and start building models, e.g.
  - Demographers
  - Economists
  - Natural scientists – climate, atmosphere, ecology, agriculture.
  - Statisticians, mathematicians
  - Computer scientists
- Data repository of relevant data
- Simulations
- Forecasts
- Communication strategy to effectively communicate to policy makers and general public

Communiation 1

Challenge

- Climate emergency and CV19 pandemic are two current examples of the breakdown of the relationship between science / scientists and the public
- If our job is to eventually create positive change in society, we must address this
- Trust is fragile or broken
- Implementing change will require public buy-in
- Social media has created new opportunities for rapid, mass dissemination of both mis- and disinformation
- There are serious coordinated efforts to discredit science

Communication 2

Action

- Not really sure what to do here, but completely convinced we need to do it
- Engage communication experts
- Include communication in training programs
- Make communication beyond article publication a required part of every project ?
- Possibly learn from corporate world how to package and sell our work successfully ?
- Continuous public engagement
- ? ? ?

About me 1

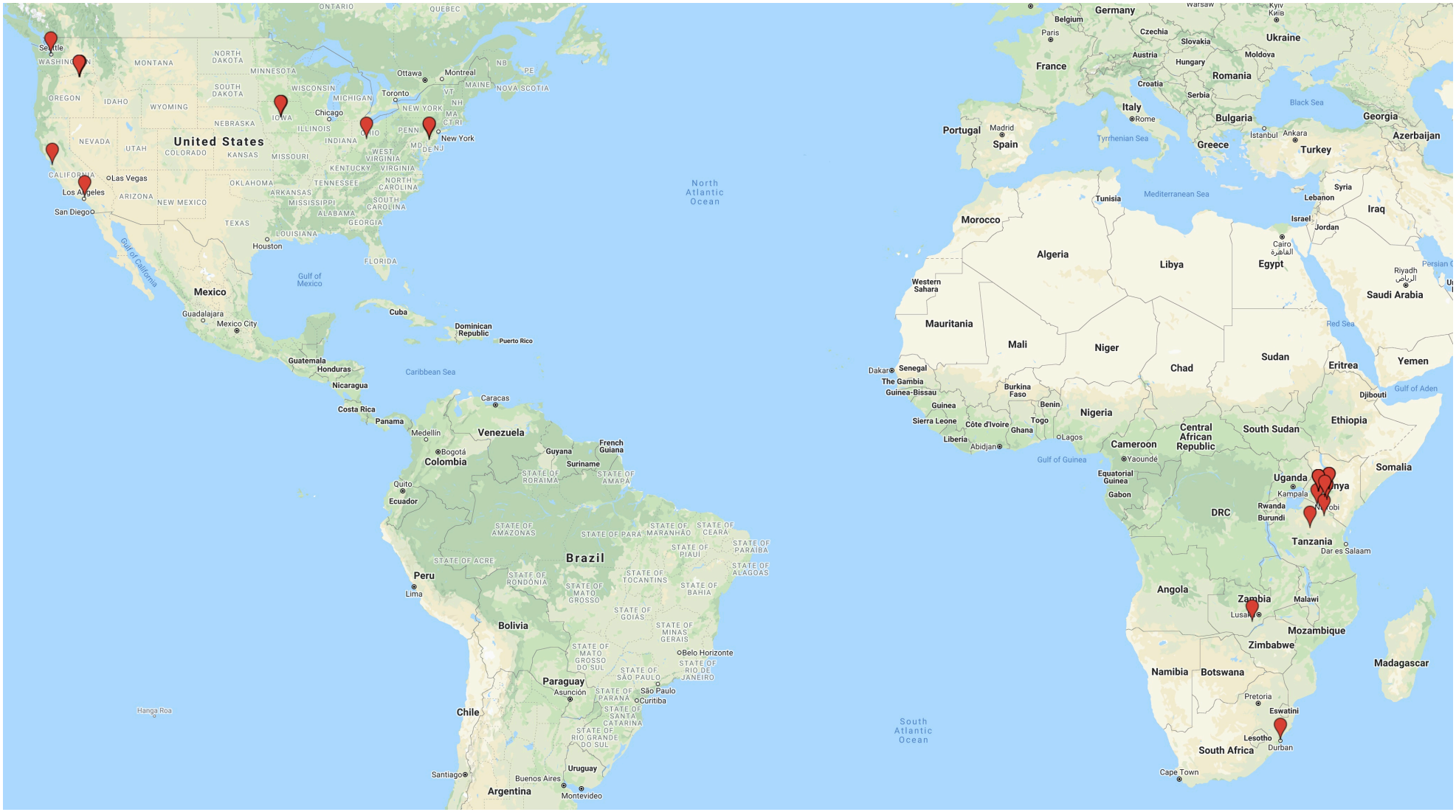
- Professor at The Ohio State University, Columbus, Ohio, USA
- Demographer / Epidemiologist / Data Scientist $\rightarrow$ multidisciplinary
- Education in biology, engineering / computer science, and demography
- Born in Kenya, grew up in East Africa, parents American $\rightarrow$ mixed identity – *not* South African!
- PhD dissertation in Zambia
- 5-year postdoc in South Africa working at demographic surveillance system sites; helped found the INDEPTH Network

About me 2

- Most of my career working on topics affecting Africa
- Current work mostly on
  - statistical / computational methods for characterizing burden of disease in areas where traditional vital statistics systems do not function
  - mathematical models of age-specific mortality
  - methods to improve coverage and accuracy of mortality estimates
- Married to Clarissa Surek-Clark who is from Brazil; hence in this country, my family members are Latinx

More information and PDFs of all publications: samclark.net, CV.

Lived/worked in East/Southern Africa and USA



23 places where Sam has lived

Research themes and applications

- Africa-related demography and epidemiology
- Orphan mortality
- Methods
- Data methods
- Interviewer effects
- Small-area estimates
- Indirect estimates of mortality
- COVID-19 in Ohio
- Verbal autopsy
- Software
- Verbal autopsy implementation

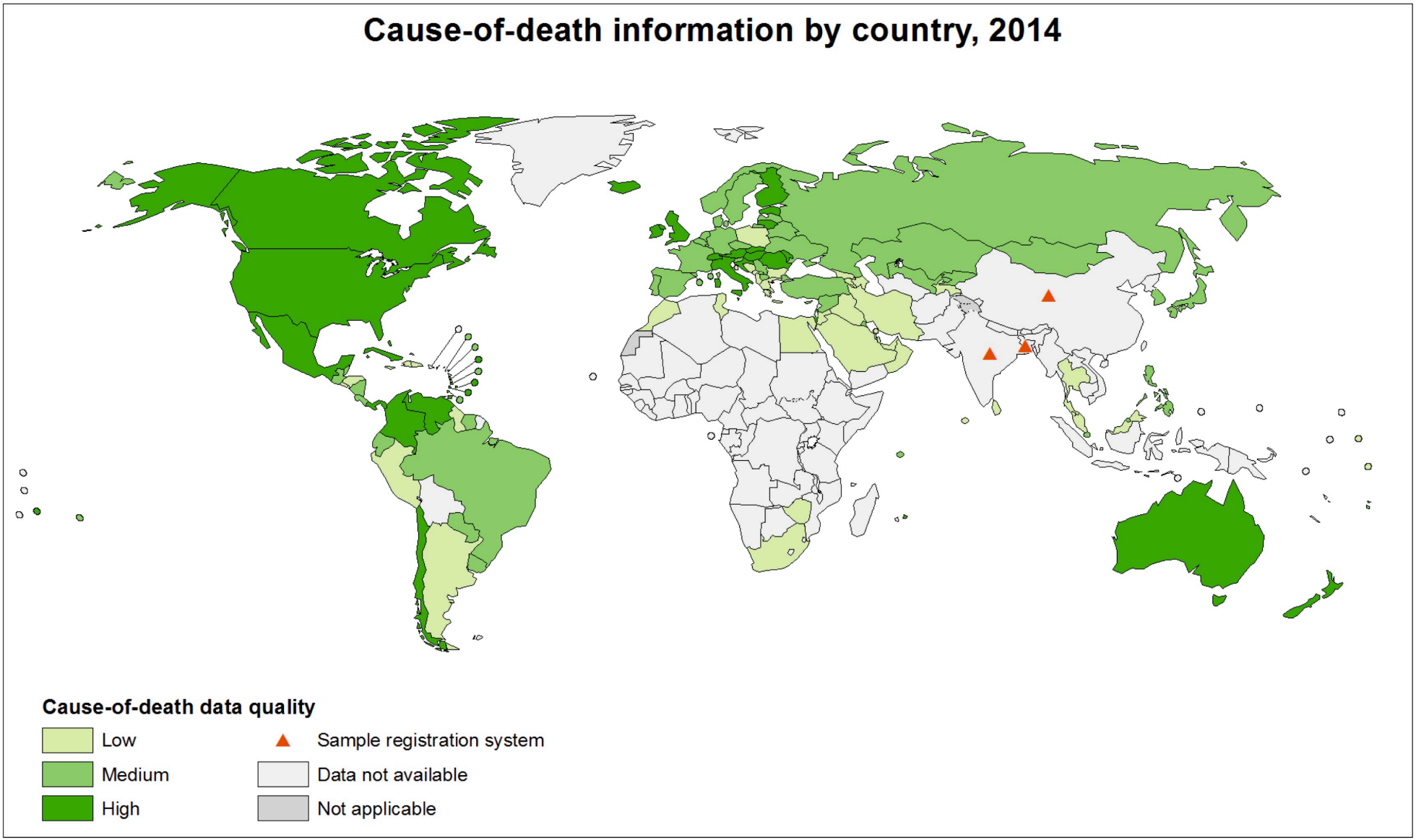
Verbal autopsy 1

Cause of death

- Essential for understanding population health, gauging effect of shocks and interventions, planning
- Difficult topic – hard to define exactly, layers of legal implications – death certificate, insurance, crimes, public health, lots of measurement error/uncertainty
- Operationally (very) difficult and (extremely) expensive
- Well-functioning vital statistics systems use autopsy, medical certification, police/coroners and others to ascertain cause $\rightarrow$ complex, expensive, requires highly-trained people
- Data systems combine several or many sources and different types of data – USA is decentralized and challenging
- Lower- and middle-income countries often have partial coverage, partially functioning systems
- Effectively all traditional means of ascertaining cause of death unavailable at scale in lower- and middle-income countries

Global cause of death

(Nichols et al., 2018)



Verbal autopsy 2

Cause of death in routine mortality surveillance

- Verbal autopsy is the only feasible (cost, complexity) means of ascertaining cause of death at scale in lower- and middle-income countries
- Nonetheless, verbal autopsy is hard to do, especially at scale

Verbal autopsy

- Interview with nearest caregiver of decedent
- Identify likely causes of death using only data from interview – no biomarkers, pathology, etc.
- Two approaches to classification: physician-review ($\approx$ differential diagnosis), or computer coding
- At scale, only computer coding is feasible

openVA

I lead the openVA project that develops methods and software tools for verbal autopsy

Global approach to improving burden of disease measures

1. Biggest need is cause of death in lower- and middle-income countries
2. Focus on verbal autopsy
3. Work with WHO to improve and develop global standards for the method: *necessary for wide adoption* (e.g. Chandramohan et al., 2021; Nichols et al., 2018)
4. Develop more accurate, efficient, and trustworthy computational/statistical methods for automated computer coding of cause (e.g. Li et al., 2020, 2019; McCormick et al., 2016; Kunihama et al., 2020)

openVA

Global approach to improving burden of disease measures

5. Improve/create trust/confidence among users
   - fractious academic literature creates confusion and does not engender public trust!
   - create open source software for **all** computer coding methods (not just ours)
   - See openVA software and openVA software Github site
6. Work with implementation partners to integrate verbal autopsy and automated cause coding into existing and developing routine mortality surveillance at national scale in many countries
   - See openVA pipeline
7. Next phase: work with funders, WHO, implementation partners to develop global repository of reference deaths at WHO to greatly improve the potential to develop new and better computer coding algorithms
8. Keep iterating ... persistence

CV19 in Ohio 1

Idea: rapidly develop CV19 public health information for use in state-level response and monitoring

State-representative CV19 seroprevalence

- 'Close' collaboration with state Department of Health (contract) aimed at immediate impact
- Seroprevalence study (Turner et al., 2021; Clark and Norris Turner, 2021)
  - De novo sampling design of adults in state
  - **Very** difficult fieldwork
  - Less than 20% response, likely along socio-political dimensions
  - Data defects required new method to effectively estimate prevalence (Kline et al., 2021)
  - Effectively ignored by state and public
- Respondent response rate and overall response to study problematic

CV19 in Ohio 2

Excess deaths study

- Real-time access to raw State of Ohio vital statistics stream
- Developed new statistical nowcasting/forecasting model for county-level excess deaths by time
- Observed real-time degradation of Ohio vital statistics system
- Pointedly ignored by state, under NDA so no publishing

Overall OH CV19 work: complete failure at pragmatic level – **communication problems**

Need a completely new approach

Summary

Challenges

- Climate emergency
- Health
- Public trust in science

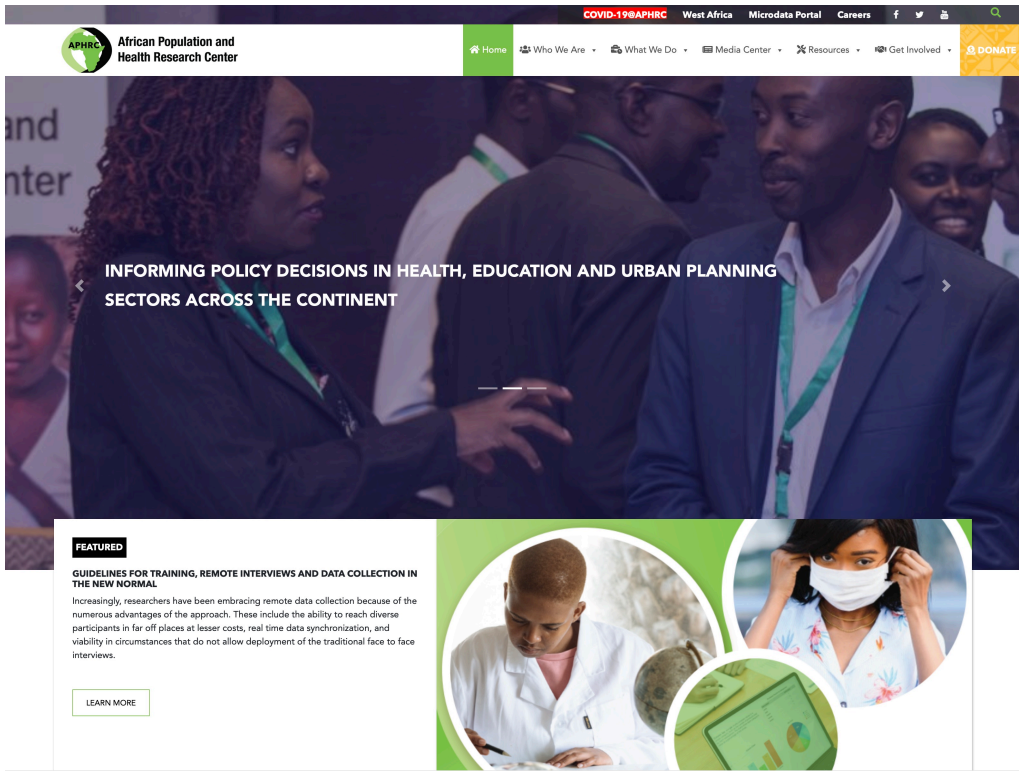
Vision

1. Develop global approach with a focus on Africa
2. Statistical demography and epidemiology
3. Environment – Population – Economy
4. Communication: population science and the public

African Population and Health Research Center

APHRC

- **Africa-led**
- Nairobi, Kenya
- Dakar, Senegal
- population & health research
- research capacity strengthening – *training*
- policy engagement and communications
- high impact, growing quickly
- URL: aphrc.org

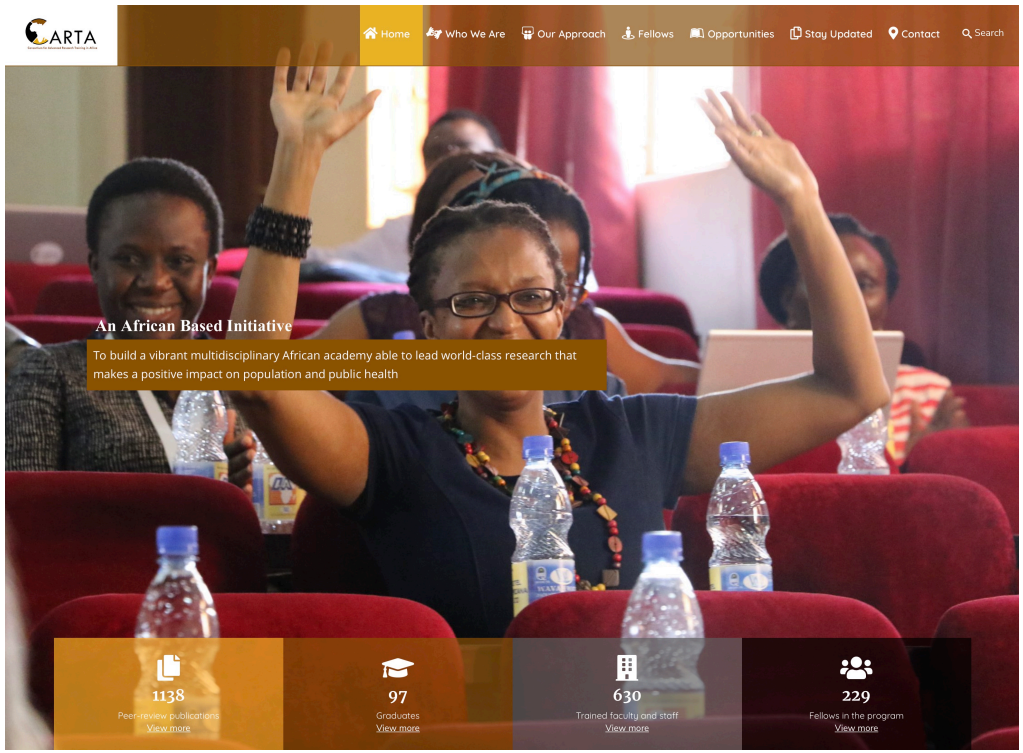


Consortium for Advanced Research Training in Africa

Mission: Build high-level capacity for population and public health-related research in Africa

CARTA

- **wholly within Africa and Africa-led**
- consortium of African universities
- PhD training across consortium
- support young researchers: PhD, postdoc, faculty
- build critical mass of locally-trained and highly effective researchers
- URL: cartafrica.org

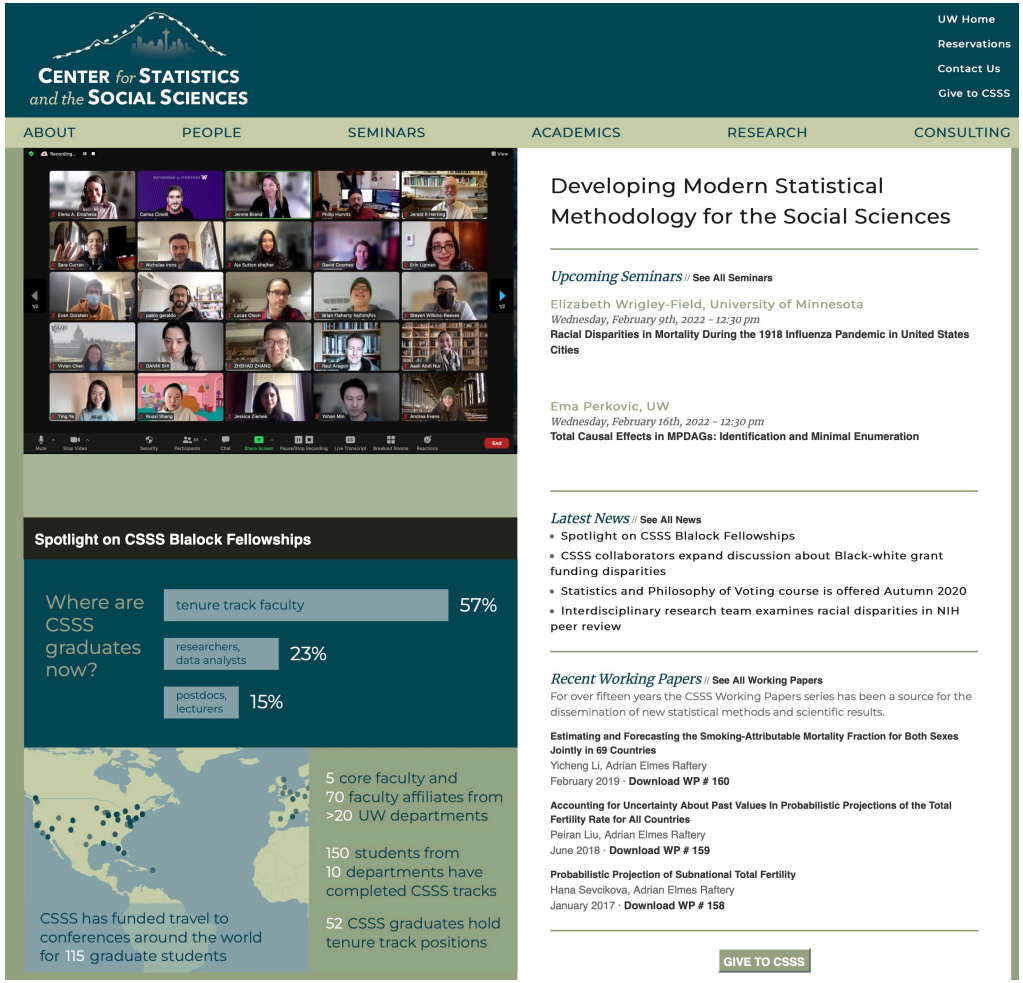


Center for Statistics and the Social Sciences

Mission: Developing Modern Statistical Methodology for the Social Sciences

CSSS, University of Washington

- bridge statistics and social sciences – core faculty from statistics and social science disciplines
- advanced statistics curriculum for Masters and PhD
- working paper series
- research and teaching like regular faculty
- URL: csss.uw.edu

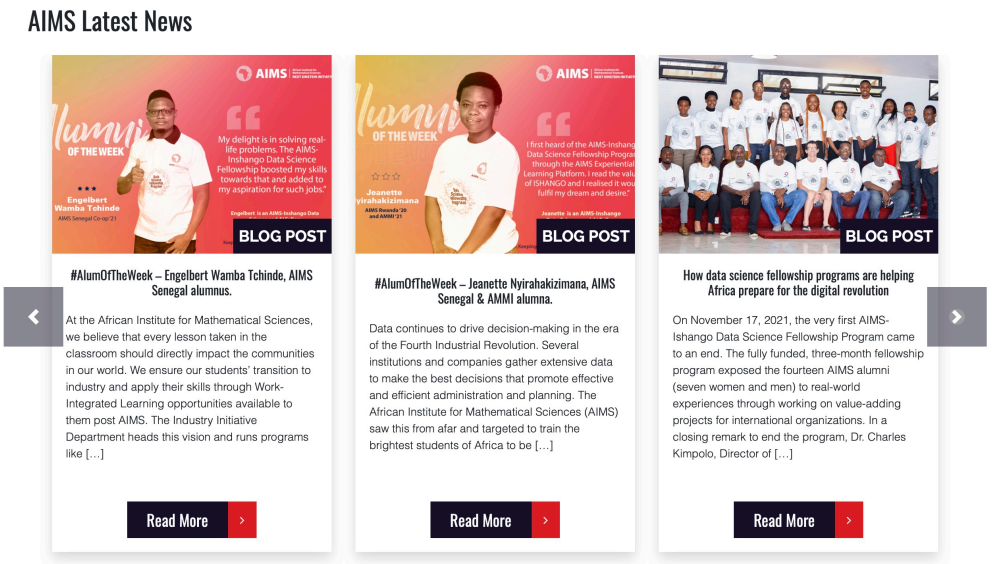


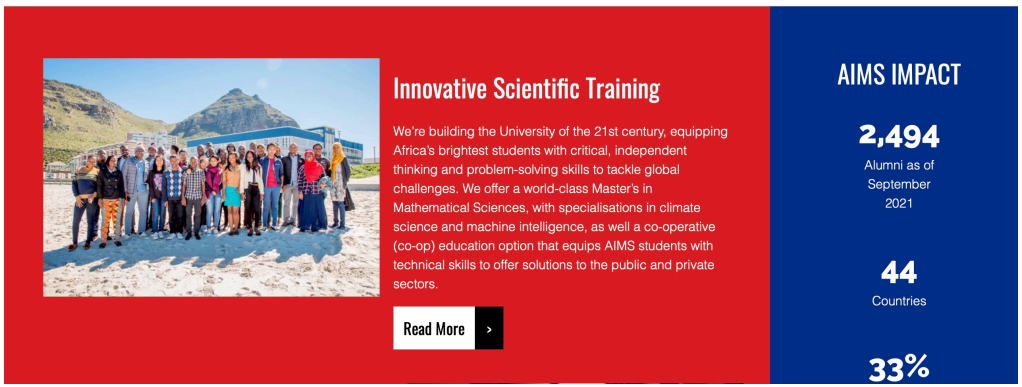
African Institute for Mathematical Sciences

Pan-African network of centres of excellence enabling Africa's talented students to become innovators driving the continent's scientific, educational and economic self-sufficiency

AIMS

- **wholly within Africa and Africa-led**
- consortium: South Africa, Cameroon, Senegal, Ghana, Rwanda
- advanced training and research
- public engagement
- cultivating STEM education in Africa
- URL: nexteinstein.org





References I

Chandramohan, D., E. Fottrell, J. Leitao, E. Nichols, S. J. Clark, C. Alsokhn, D. Munoz, C. AbouZahr, A. Di Pasquale, R. Mswia, E. Choi, F. Baiden, J. Thomas, I. Lyatuu, Z. Li, P. Larbi-Debrah, Y. Chu, S. Cheburet, O. Sankoh, A. Badr, D. Fat, P. Setel, R. Jakob, and D. de Savigny (2021). Estimating causes of death where there is no medical certification: Evolution and state of the art of verbal autopsy. *Global Health Action In press*.

Clark, S. J. and A. Norris Turner (2021, March). *invited commentary*: monitoring epidemics: lessons from measuring population prevalence of the coronavirus. *Proceedings of the National Academy of Sciences 118*(9), e2026412118.

Kline, D., Z. Li, Y. Chu, J. Wakefield, W. C. Miller, A. N. Turner, and S. J. Clark (2021). Estimating seroprevalence of SARS-CoV-2 in Ohio: A Bayesian multilevel poststratification approach with multiple diagnostic tests. *Proceedings of the National Academy of Sciences 118*(26).

Kunihama, T., Z. R. Li, S. J. Clark, and T. H. McCormick (2020). Bayesian Factor Models for Probabilistic Cause of Death Assessment with Verbal Autopsies. *The Annals of Applied Statistics (Preprint: arXiv:1803.01327v2) 14*(1), 241–256.

Li, Z. R., T. H. McCormick, and S. J. Clark (2019, June). Bayesian Joint Spike-and-Slab Graphical Lasso. In *Proceedings of the 36th International Conference on Machine Learning*, Volume 97 of *Proceedings of Machine Learning Research*, pp. 3877–3885. International Conference on Machine Learning, Long Beach, California, USA.

References II

Li, Z. R., T. H. McCormick, and S. J. Clark (2020). Using Bayesian Latent Gaussian Graphical Models to Infer Symptom Associations in Verbal Autopsies. *Bayesian Analysis 15*(3), 781–807.

McCormick, T. H., Z. R. Li, C. Calvert, A. C. Crampin, K. Kahn, and S. J. Clark (2016). Probabilistic Cause-of-Death Assignment using Verbal Autopsies. *Journal of the American Statistical Association 111*(515), 1036–1049.

Mikkelsen, L., D. E. Phillips, C. AbouZahr, P. W. Setel, D. De Savigny, R. Lozano, and A. D. Lopez (2015). A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *The Lancet 386*(10001), 1395–1406.

Nichols, E. K., P. Byass, D. Chandramohan, S. J. Clark, A. D. Flaxman, R. Jakob, J. Leitao, N. Maire, C. Rao, I. Riley, et al. (2018). The who 2016 verbal autopsy instrument: an international standard suitable for automated analysis by interva, insilicova, and tariff 2.0. *PLoS medicine 15*(1), e1002486.

Nichols, E. K., P. Byass, D. Chandramohan, S. J. Clark, A. D. Flaxman, R. Jakob, J. Leitao, N. Maire, C. Rao, I. Riley, and P. Setel (2018). The WHO 2016 Verbal Autopsy Instrument: An International Standard Suitable for Automated Analysis by InterVA, InsilicoVA, and Tariff 2.0. *PLoS Medicine 15*(1), e1002486.

Turner, A. N., D. Kline, A. Norris, W. G. Phillips, E. Root, J. Wakefield, Z. R. Li, S. Lemeshow, M. Spahnie, A. Luff, Y. Chu, M. K. Francis, M. Gallo, P. Chakraborty, M. Lindstrom, G. Lozanski, W. Miller, and S. J. Clark (2021). Prevalence of current and past covid-19 in ohio adults. *Annals of epidemiology*.

References III

UN DESA Population Division (2019). World population prospects 2019. https://population.un.org/wpp/ and https://github.com/sinafala/World-Population-Fractions.