Multi-disciplinary Data Science to Improve Global Health: Verbal Autopsy

Sam Clark

Notestein Seminar
OPR, Princeton
September 12, 2023

Data science for social good - DSSG

- Social
  - Something related to the welfare of individual people and/or groups of people
- Good
  - Improvements can be made (and?) commercial/market-driven system not acting
  - *Something specific* has to change in a positive way
  - Change needs to be *sustained*
- Data science
  - New information can help
  - New analytical approaches can help
- Research
  - Create new methods that incorporate (new?) data and result in new knowledge
- Implementation
  - *Translate research into sustained, useful practice*

Questions about DSSG

- Academia
  - Institutions talk a lot about it, claim to value it, few specifics!
  - How should academics train and be involved?
- *How do we keep analytics honest, fair, representative?*
- Who else is involved and how do we work effectively with them?
- What is ‘research’ in the context of DSSG
  - How do we get academic credit - e.g. credit toward tenure/promotion - for this work?
  - What are products or impacts that can go on an academic CV?
- Potentially a lot of money involved
  - Fund raising: foundations, industry, etc.?
  - Salable products, start-ups, for-profit partnerships?
- What is ‘data science exactly’ - data, methods, and ?
- What responsibility do we have to apply results *legally*, fairly, usefully

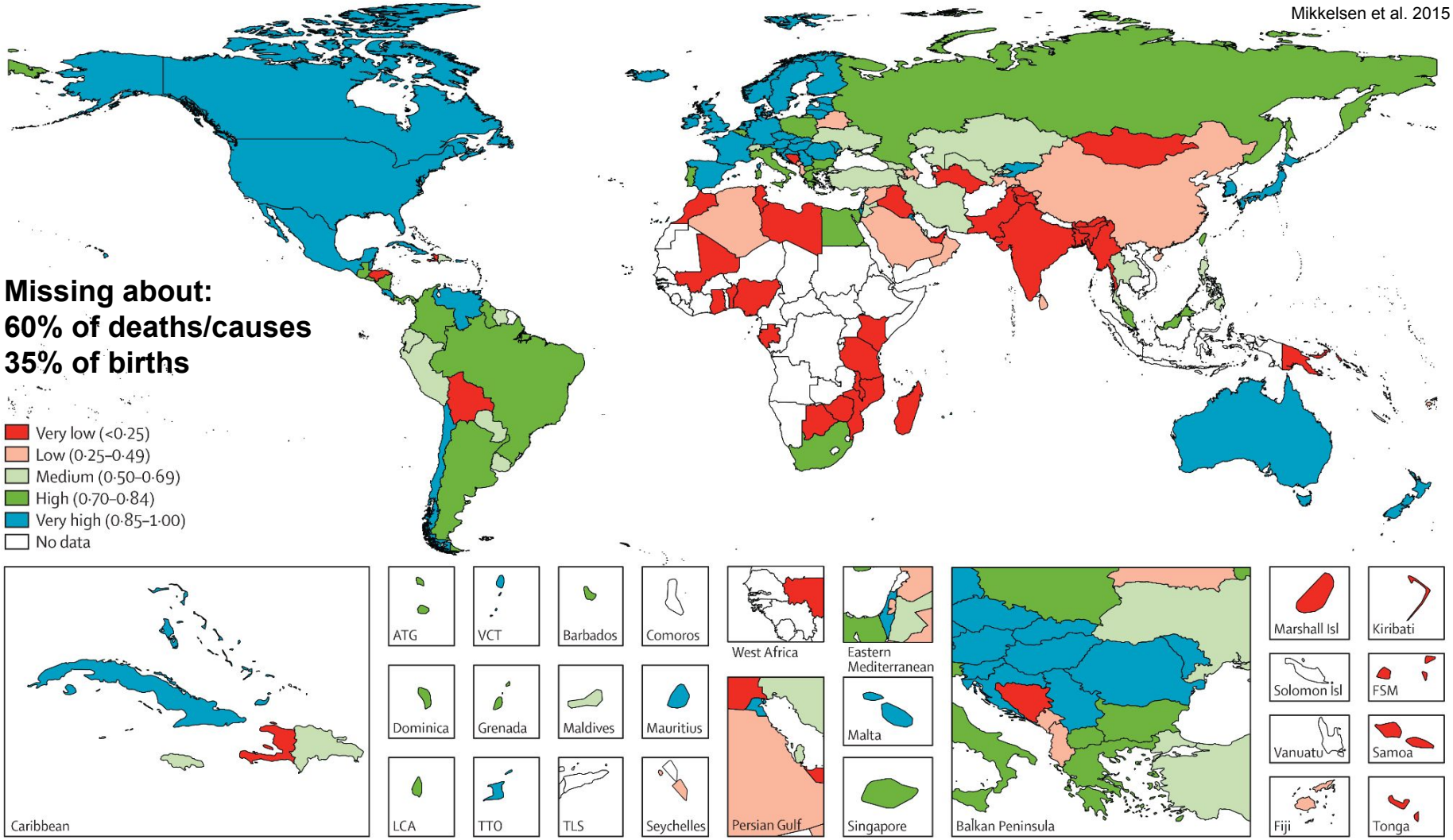
Guidelines & accountability for academics doing DSSG

1. Is the ultimate objective a consequential social good?
2. Is the approach
    a. Affecting inequality in a negative way?
    b. Feasible in the target context?
    c. Self-sustainable in the target context?
3. What are potential unintended consequences, positive and negative?
4. Can negative unintended consequences be resolved?
5. Is there a feasible, useful plan for the future
    a. How do academics disengage?
    b. If academics must be continually involved, how is this supported?
6. Is the overall approach academically sustainable
    a. Academic training
    b. Academic career progression: hiring, tenure, promotion

How did I get to DSSG?

- Born in ***Kenya*** and and grew up ***Tanzania*** and ***Kenya***
- Moved to USA at age 16
- Undergraduate at Caltech
  - BS Biology (neurobiology)
  - BS Engineering (computer science, electrical engineering)
- Graduate work at Penn
  - PhD Demography
  - Fulbright field work in Zambia
- Postdoc in South Africa
  - 5 years working with health and demographic surveillance system sites
- About 10 years working with statisticians and UN Populatino Division on methods

Mikkelsen et al. 2015



Verbal Autopsy

Verbal autopsy basics

- Old method - 50s onward (Chandromohan et. al, 2021)
- *Only* feasible method where traditional death certification/autopsy are not possible
- *Very* challenging in general
  - Interview is difficult
  - Data are inherently noisy and contain comparatively little information
  - Cause classification task has many dimensions compared to information
  - *Cause classification is extremely challenging!*
  - Results are noisy and uncertain and require careful interpretation
  - Although feasible, implementation still very difficult

- **Consequential, many potential users, and lots to do 🙂**

Verbal autopsy basics

- Predefined
  - *VA causes*
  - Important signs/symptoms - *VA indicators*
  - *Symptom-cause information* that links indicators with causes
- Interview close caregiver(s) of recently dead
  - 100s of close-form questions
  - Narrative account of death
- Data are used to classify death into VA cause list
- Result
  - **COD**: individual-level cause of death in predefined VA cause list
  - **BOD**: population-level cause-specific mortality fractions (CSMF) - distribution of deaths by VA cause

VA instrument

- Two standard instruments
- WHO global standard
  - Early 2000s until now
  - Number of versions and revisions
- Population Health Metrics Research Consortium (PHMRC) standard
  - Mid 2000s until now
  - ‘Long’ and ‘short’ versions
- Each instrument family is foundation of it’s own ecosystem of
  - Interview
  - Coding mechanisms
  - Interpretation

VA Interview

- Standard instruments in electronic form
  - WHO uses Open Data Kit (ODK) definition
- Conducted in many languages and socio-cultural contexts
- Apart from standard training materials, not much done to standardize
- Implementers always modify and add/subtract
- Very complex, long interview
- Emotionally challenging/draining for both respondent and interviewer
- Lots of different languages and cognitive frameworks involved across resource-constrained settings where VA used
- *Important source of consequential variability in VA data*
- Urgent area for more research and improvement

VA data

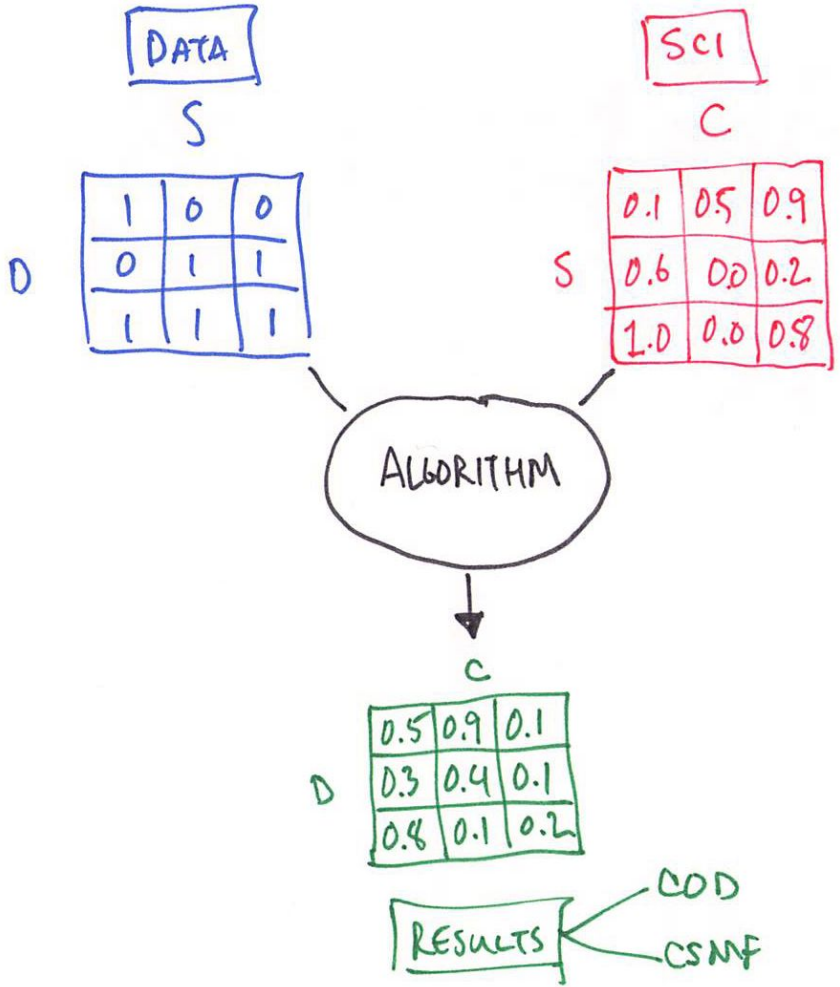
- Complex instrument and challenging interview → require systematic checking for quality and consistency
- Two forms of VA data
- Raw data
  - Full responses to closed-form questions
  - *Sometimes* narrative account as free-form text
- Processed VA indicators for automated algorithms
  - Closed-form questions converted to algorithm-ready binary indicators
  - Narrative account processed to produce word counts
  - Narrative account as free text

Physician-coded VA - PCVA

- Traditional approach to classifying VA deaths by VA cause
- Physicians read interview transcript and identify most likely cause(s)
- Physicians often disagree
- Standard practice
  - Multiple physicians code
  - Consensus process to produce final cause
- CSMFs as sum of deaths with each cause
- Characteristics
  - Comparatively specific causes
  - More accurate in general
  - Non-replicable, inconsistent, contextual
  - Slow, costly
  - Takes often scarce physician time from living patients

Computer-coded VA - CCVA

- 3 components
  - VA data
  - Symptom-cause information (SCI)
  - Algorithm logic
- Characteristics
  - Cheap, fast
  - Replicable
  - Does not require physician time
  - Can be fully automated
  - Can be integrated into routine systems
  - Comparable
  - Comparatively less specific causes
  - Less accurate in general
  - Challenges with context-specific performance



VA cause of death - COD

- Individual-level cause of death using VA cause list
- Physicians produce either single cause or underlying, immediate, contributing
- Algorithms
  - Produce a decimal-valued metric for each cause, large values more likely
  - Metrics used to rank causes, usually cause with largest metric value is reported
  - For some algorithms, if no causes have metric above given threshold, then reported cause is ‘undetermined’
  - Most algorithms are deterministic and do not report uncertainty/confidence
  - **InSilicoVA** is probabilistic and produces distributions of probabilities for each cause

VA burden of disease - cause-specific mortality fractions

- Burden of disease (BOD) is distribution of deaths by cause
- Population-level metric consisting of distribution of deaths by cause for a specified population
- VA terminology - cause-specific mortality fractions (CSMF)
- In most circumstances - physicians and most algorithms
  - Top cause for individual deaths summed to create CSMFs
  - Deterministic process that does not translate any uncertainty in individual COD to population level
  - **InSilicoVA** produces distributions of CSMFs for all causes that are consistent with underlying individual-level COD probability distributions

My goals for VA work

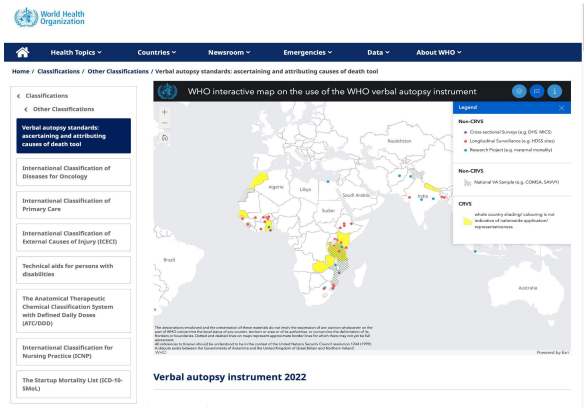
- Research
  - Create and maintain/evolve global standards for VA
  - Greatly improve replicability and comparability of VA COD, BOD - VA results comparable from regions, historical periods, specific contexts
  - Improve VA interview → greatly improve VA data
  - Improve VA algorithms - accuracy, uncertainty, context-specific performance, broad comparability of results, account for changing epidemiological circumstances over time, etc.
  - Create robust software - research and routine use
  - Create Reference Death Archive - SCI, algorithm development and validation
- Implementation
  - Create methods and software to integrate automated VA algorithms into routine mortality surveillance
  - Support roll-out of routine-use VA software - CRVS
  - Continuous reporting of COD and BOD at scale in wide variety of settings

The openVA Team

Implementing Countries:

- Rwanda
- Ethiopia
- Tanzania
- Zambia
- Solomon Islands
- About 20 more waiting

AHRI AFRICA HEALTH RESEARCH INSTITUTE

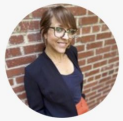


MEDICINA USP

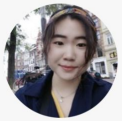
openVA Team

Research Team

Sam Clark
Principal Investigator



Nicole Angotti

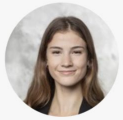


Yoonyoung Choi

Sherry Zhao

Eungang Peter Choi

Yue Chu



Melina Raglin

Jason Thomas

Tyler McCormick

Zehang Richard Li

Clarissa Surek-Clark

Zhenke Wu

Supporters & Partners

NIH NICHD

NICHD

ALPHA NETWORK

Alpha Network

Vital Strategies

Vital Strategies

CDC Foundation
Together our impact is greater

CDC Foundation

The Ohio State University

The Ohio State University

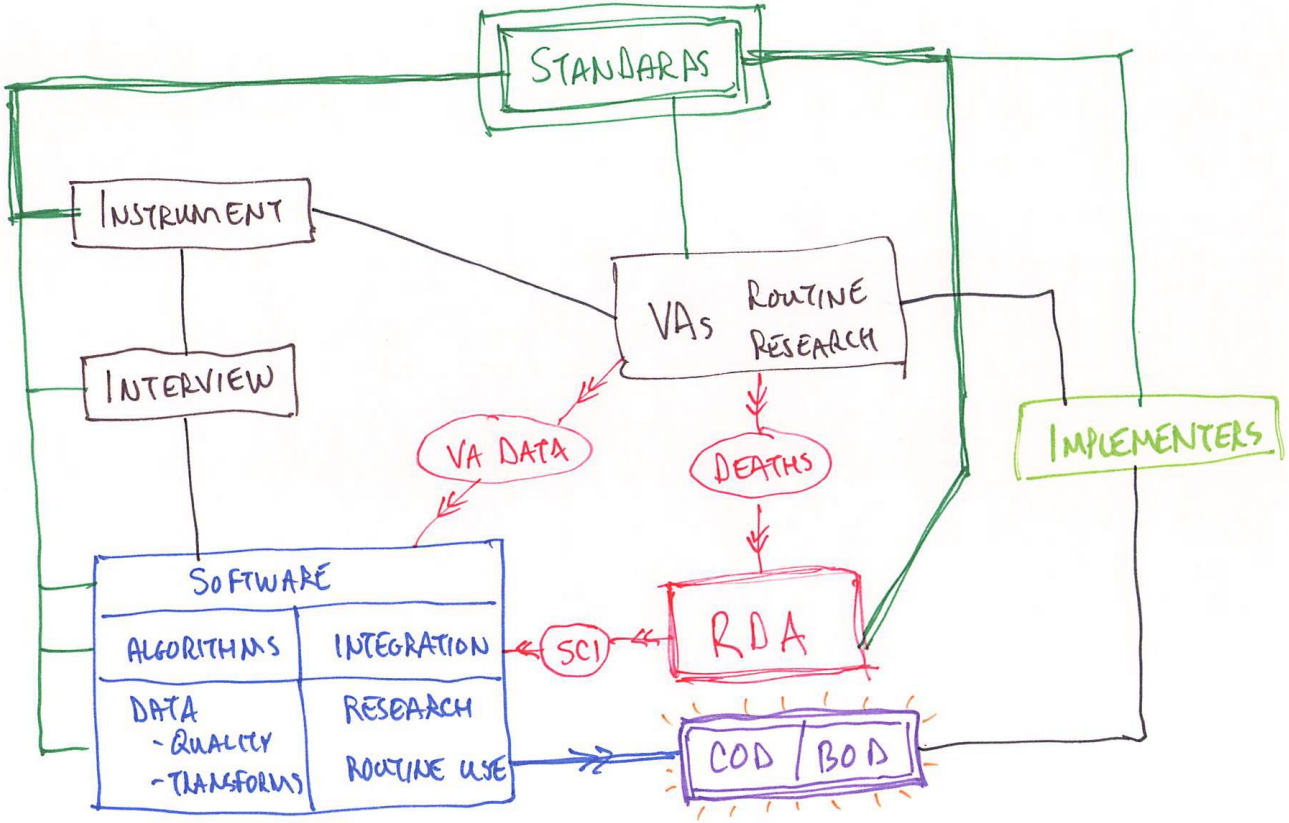
Bloomberg Philanthropies
DATA FOR HEALTH INITIATIVE

Bloomberg Philanthropies

U.S. Centers for Disease Control and Prevention

CDC

What the openVA Team does



VA Interview

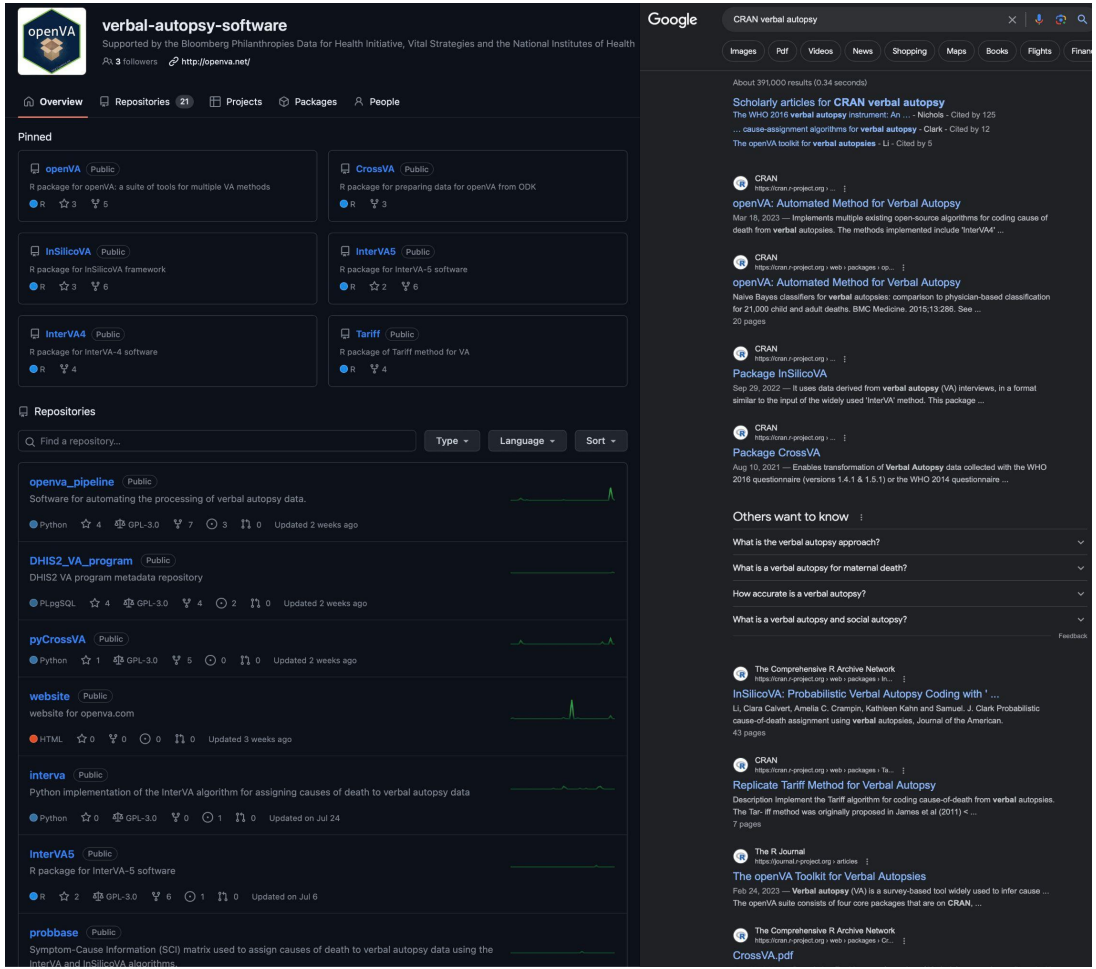
- Objective: standardize VA interview to extent possible
  - Improve data quality
  - Improve comparability of VA overall
- Challenges
  - Wide variety of languages, cognitive frameworks
  - Poor training and maintenance of VA field teams
  - Extremely heterogeneous interview design, no standards, effectively no training
  - Electronic data capture is poor quality, uneven, affects data: ODK
- Observe VA field operation and interview in many settings
- Characterize VA interview language usage and translation practices
- Address issues!

VA cause-ascertainment algorithms - 1

- Extremely fractious field/literature
  - Disingenuous publishing
  - Rivalries
  - *Everything hidden, proprietary, non-reproducible*
- Replicate/reverse engineer all algorithms
- Create open-source versions of all algorithms
- Create standard software framework for running all algorithms
- Create new algorithm - **InSilicoVA**
- Work to improve symptom-cause information for all algorithms
- Work on next generation of algorithms
  - Incorporate dependence among VA indicators
  - Address domain specificity
  - Improve computational performance

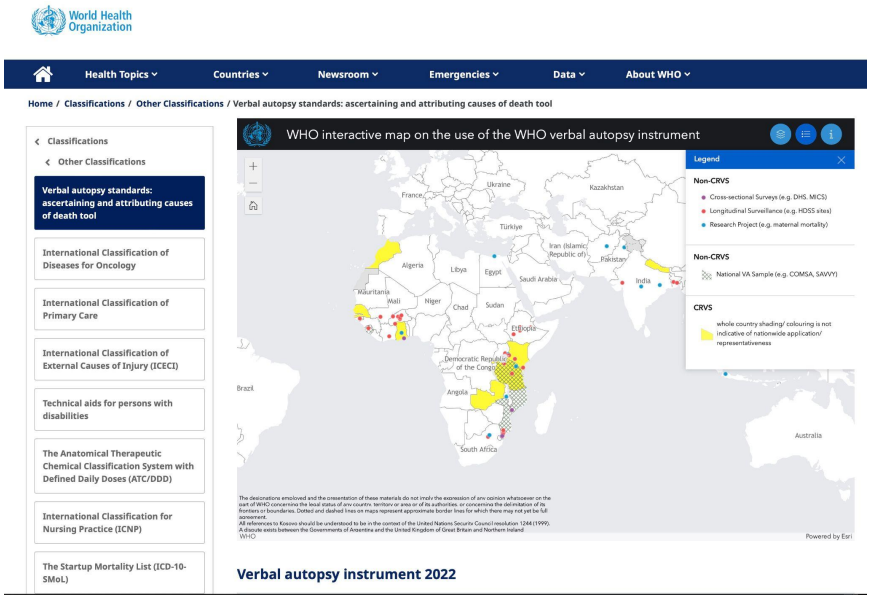
VA software

- Very firm commitment to open source software paradigm → **openVA**
- All theoretical work on algorithms accompanied by complete, working, open-source code
- Unify software application of all algorithms
- Maintained R packages on CRAN
- Maintained Python code
- Github repos for all



Global VA standards

- Member of WHO Verbal Autopsy Reference Group (VARG)
- Create and maintain WHO standard VA instrument
- ODK definition for instrument
- Training materials for VA field team
- Question-by-question guidance for interviewers
- Wide variety of supporting tasks
- [VARG web page]



Reference Death Archive - 1

- Algorithms combine *VA data* and *symptom-cause information* - **SCI** - to classify deaths by cause
- SCI describe relationships between VA algorithm indicators and VA causes
- *SCI drive operation of algorithms*
- For CCVA to be accurate and comparable
  - SCI must reflect changing epidemiological circumstances and localized nuances in indicator/cause relationships
  - Must be updated through time to reflect changes
  - Must be widely available in standard form appropriate for algorithms
- Sources for SCI
  - Crowdsource from physicians: Pr(s|c)
  - *Calculate directly from* ***reference deaths with both standard VA and reference cause***

Reference Death Archive - 2

- Crowdsourced SCI comparatively low information
- Because of dimensionality and unevenness of VA classification task, reference death SCI require large number of deaths: 10^5 - 10^6
  - Deaths not evenly distributed by cause
  - Roughly 10,000-15,000 meaningful relationships
  - Need reasonable number of observations for each relationship to characterize conditionally independent relationships and dependence - big challenge!
- To cover variety of epidemiological/localized circumstances, reference deaths must be drawn from wide variety of settings
- To follow temporal evolution of relationships, reference deaths must be continuously contributed

Reference Death Archive - 3

- Build an archive of reference deaths
- Hosted by reliable, trustworthy 3rd party: WHO
- Available to all VA researchers and VA users
- Researchers
  - Create/improve SCI
  - Develop/improve algorithms
  - Test/validate algorithms
- VA users have access to continuously updated SCI
  - For all algorithms
  - Cover wide variety of epidemiological circumstances and time
  - Allow comparable CCVA cause
  - Archived for future reference
  - Integrated with CCVA software - ease of use

What the openVA Team has accomplished so far

VA Interview

- Very challenging situation!
- Observed interview in many diverse socio-cultural, linguistic settings
- Characterized variation in and developing recommendations to partially standardize and improve
  - Interview design
  - Language use
  - Translation and interpretation
- Documented different approaches to eliciting narrative account
- Working to develop algorithm-assisted ‘dynamic interview’
- Observed emotional strain on interview/field teams
- Developing recommendations to improve emotional health of all involved

VA algorithms

- Created, tested, validated, and many applications of InSilicoVA Algorithm
  - Individual and population-level consistent probabilistic estimates of COD and CSMF
  - Consistent uncertainty in COD and CSMF
  - No ‘undetermined’ causes - hard to classify deaths are very uncertain - flat, wide distribution
  - Reasonable computational performance
- Re-implemented all existing algorithms except Tariff 2.0 in R and Python
- Developed a number of approaches to addressing dependance among symptoms, e.g. https://arxiv.org/abs/2112.12186 (accepted to AOAS yesterday)
  - For now, computationally too expensive for routine use
  - Need better (empirical) SCI
- Conducted only fully fair comparison of algorithm performance
  - Accounted for variation in SCI and algorithmic logic → SCI more important
  - No algorithms perform extremely well; InSilicoVA performs best all together but not in every single situation
- Developing new approach to incorporate augmented SCI - MITS

VA software

- **openVA** suit of algorithms in R and Python
- **openVA Pipeline** for CRVS integration - fully automates:
  - Retrieval of VA deaths from ODK Connect server
  - CCVA cause assignment using any algorithm
  - Packaging raw data and results of assignment into compact format
  - Push full results to DHIS-2 (District Health Information System)
- crossVA in R and Python
  - Data checks and transformations
- Various methods of delivering workable software: packages, Docker, executable
- openVA Toolkit publication in *R Journal* as companion to software
- User-friendly GUI openVA Application in Python (new, not on PyPI yet)
  - What CRVS users actually want right now
- Open source packages and modules on CRAN, Github, and PyPI

Reference Death Archive

- New project starting now with support from Bill and Melinda Gates Foundation
- Reference deaths with WHO standard VA, independent reference cause, and *minimally-invasive tissue sample (MITS) information*
- Building computational infrastructure to be hosted by WHO in Geneva
- Sourcing reference deaths from
  - CHAMPS and COMSA mortality surveillance projects
  - Brazilian vital statistics system, starting in Sao Paulo
- Future deaths sourced from same places plus wide variety of VA users, including routine mortality surveillance at national scale/CRVS
- Planned products
  - Reference deaths easily available in well-documented and standard form via secure enclave
  - Regularly updated SCI for all algorithms available through API

Limitations

- VA is still a low-information, extremely challenging method - accuracy stubbornly low
- No current method fully accounts for uncertainty in interview/data, SCI, and algorithm - all results are over-confident, only InSilicoVA attempts this
- Communication of limitations to end users almost non-existent - urgent need to develop training for decision/policy makers on how to interpret, utilize VA results
- All current algorithms assume conditional independence of symptoms given cause - addressing this will greatly improve accuracy and uncertainty, requires much better SCI and clever implementation to make computation feasible
- Interview needs attention in many ways - standardization, streamlining, emotional strain, etc.
- …

How are we doing on the DSSG guidelines?

1. Is the ultimate objective a consequential social good? **YES**
2. Is the aproach
   a. Affecting inequality in a negative way? **MAYBE - currently, SCI not general enough**
   b. Feasible in the target context? **YES**
   c. Self-sustainable in the target context? **PROBABLY**
3. What are potential unintended consequences, positive and negative? **MANY, uncertain causes used inappropriately or with too much confidence - legal issues and overconfident public health decision-making, etc.**
4. Can negative unintended consequences be resolved? **PROBABLY, needs much more effort and focus**
5. Is there a feasible, useful plan for the future
   a. How do academics disengage? **GETTING THERE!**
   b. If academics must be continually involved, how is this supported? **SOLUTIONS EXIST**
6. Is the overall approach academically sustainable **YES**
   a. Academic training **YES**
   b. Academic career progression: hiring, tenure, promotion **YES**

Future work

- Continue to improve algorithms
  - Accuracy/uncertainty
  - Dependence
  - Domain adaptation
  - Computation performance
  - Incorporate additional information into process through augmented SCI - MITS, medical records, etc.
- Continue to create software tools for routine, large-scale use - CRVS
- Develop guidelines/recommendations to improve interview, test them
- Finish the RDA and stimulate contribution of quality reference deaths with additional information - MITS, medical records, etc.
- **Create and support a consortium of VA researchers and software developers**
- **Continue efforts to decolonize this work**
- Continue contributing to global standards and development of standard training materials

Reflections on DSSG questions - 1

- Attempting to do DSSG well is complex and difficult
- Many new kinds of partners and funders required
  - Traditional research support
  - Implementation partners and support through contracts
  - Time-consuming participation in community standards, governance - WHO
- Need to develop large, diverse network - a lot of time and effort
- Need many disciplines
  - Demography, epidemiology, statistics, computer science, linguists, anthropology
  - Can they work together effectively?
  - How to manage everyone so that it’s worthwhile in each of their contexts?
- Need diverse set of users to test and implement

Reflections on DSSG questions - 2

- Impact beyond academic publishing
  - Software, real-world impacts, advocacy, etc.
  - Goes beyond traditional academic research, teaching, academic service
  - *Can one get academic credit, get hired or promoted ?*
- **Valid efforts to create/maintain fair algorithmic methods requires massive extra, continuing, testing and validation - way beyond what's necessary for a paper**
- Many machine learning-type methods require *massive training data* - whole new challenge for academic research
  - Also where the unfairness and bias creep in
  - Need a whole lot more work on general principle for this in social, behavioral, and health sciences
- Money - tempting to make it about money, have to balance this with 'social good' - no suggestions for how to resolve this!

How to prepare for this kind of work - *my take*

- **Identify a very consequential question and corresponding application**
- Must have training in
  - **computer science** and real programming - programming a script-based, interpreted statistics package is not real programming!
  - **Statistics**
  - **Substantive field -** sociology, anthropology, demography, or epidemiology, etc.
- To run this kind of project, also need leadership training
  - **People, project, and financial management**
  - These rapidly become big projects with lots of moving parts that need significant support and competent management
- Work to develop **diverse network of colleagues** and mentors/advisors
- Be willing to do lots of things that are ‘out of your comfort zone’
- Develop a **diverse set of funders** - *research funders don’t usually fund implementation and vice versa, this requires both*
- Withstand lots of narrow critiques of what you’re doing!
- Accept slower progression through your career

Acknowledgements

- **All joint work with the openVA Team and its supporters and collaborators, see openva.net (link) - thanks to all!**
- For publications and working papers related to VA (and other things), see Sam's bibliography (link)
- We are grateful for support from
  - NICHD
  - Bloomberg Philanthropies' Data for Health Initiative
    - CDC Foundation
    - Vital Strategies
  - Bill and Melinda Gates Foundation
  - The Ohio State University

Mikkelsen et al. 2015