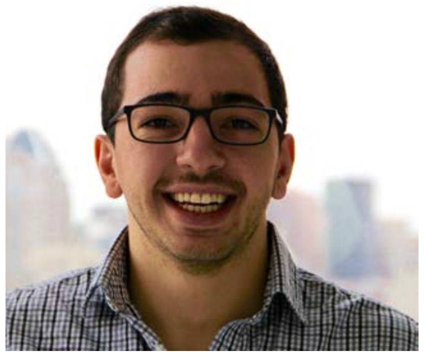
Correcting for verbal autopsy misclassification bias in cause-specific mortality estimates

Abhi Datta
Department of Biostatistics
Johns Hopkins University

Collaborators

JHU, Biostatistics

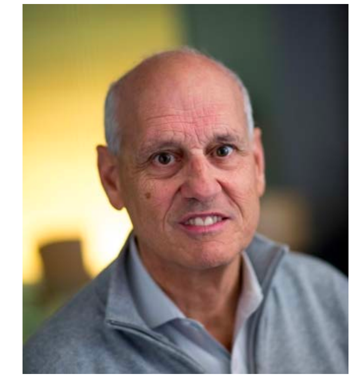


Jacob Fiksel

Brian Gilbert

(former PhD students)

Sandipan Pramanik
(Postdoctoral fellow)

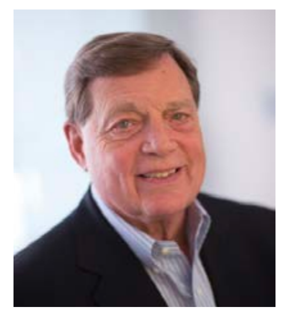


Scott Zeger

Agbessi Amouzou

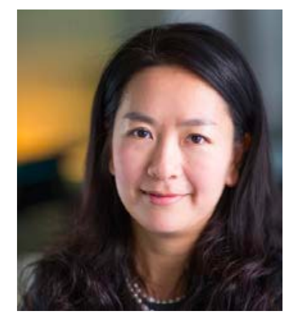
Emily Wilson

Henry Kalter



Bob Black

JHU, International Health



Li Liu
(JHU, PFRH)

Dianna Blau
(CHAMPS)

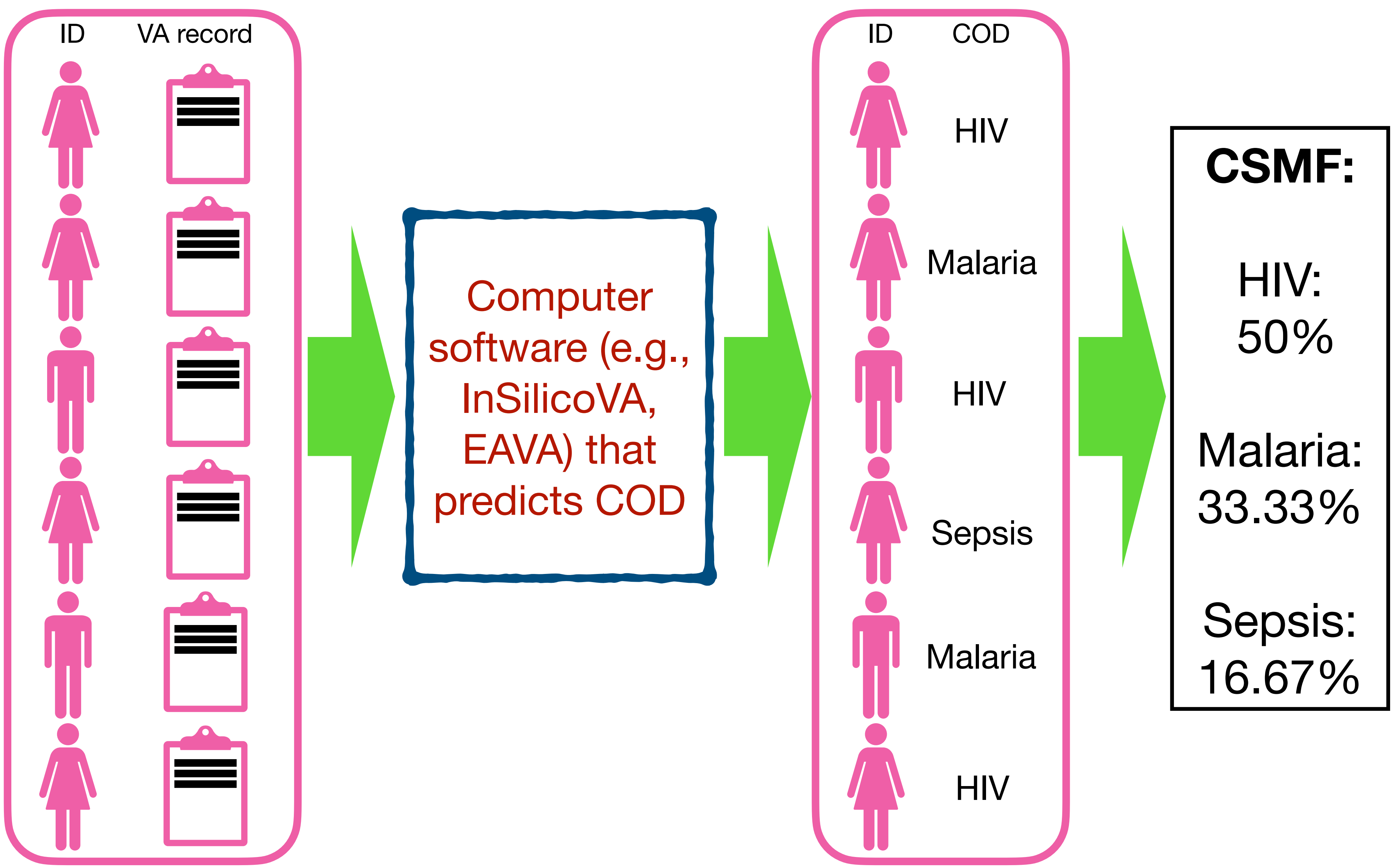
Ivalda Macicame
(INS Mozambique)

- Funding (Bill and Melinda Gates Foundation):
  - Countrywide Mortality Surveillance for Action (COMSA) in Mozambique (PI: Amouzou)
  - CA-CODE: Child and Adolescent Causes of Death Estimation (PI: Liu)
  - Broadening the applicability of minimally-invasive-tissue-sampling (MITS)-based verbal autopsy (VA) calibration to improve global mortality estimates (PI: Datta)

VA to COD to CSMF

COD = cause of death
CSMF = cause-specific mortality fractions

COMSA-Mozambique goal: Generate publicly available data on CSMFs at national and subnational levels using VA-records

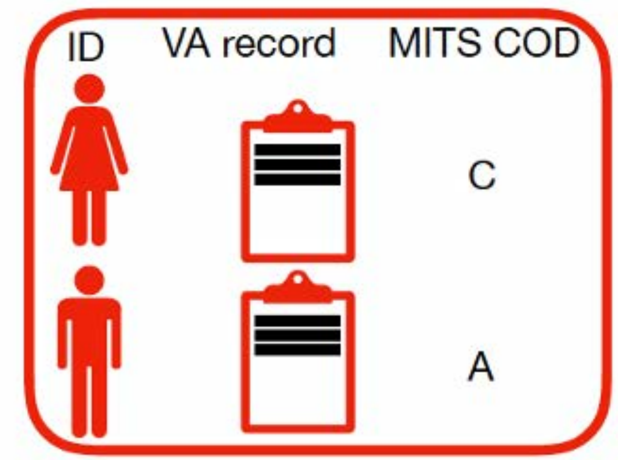


Minimally Invasive Tissue Sampling (MITS)

Data from CHAMPS project with both VA-COD and a minimally invasive tissue sampling (MITS)-COD

MITS-COD assignments been shown to be reasonably accurate when compared to the full diagnostic autopsies (Bassat et al. 2017)

CHAMPS data can be used to create a paired VA-MITS dataset to understand the accuracy of VA



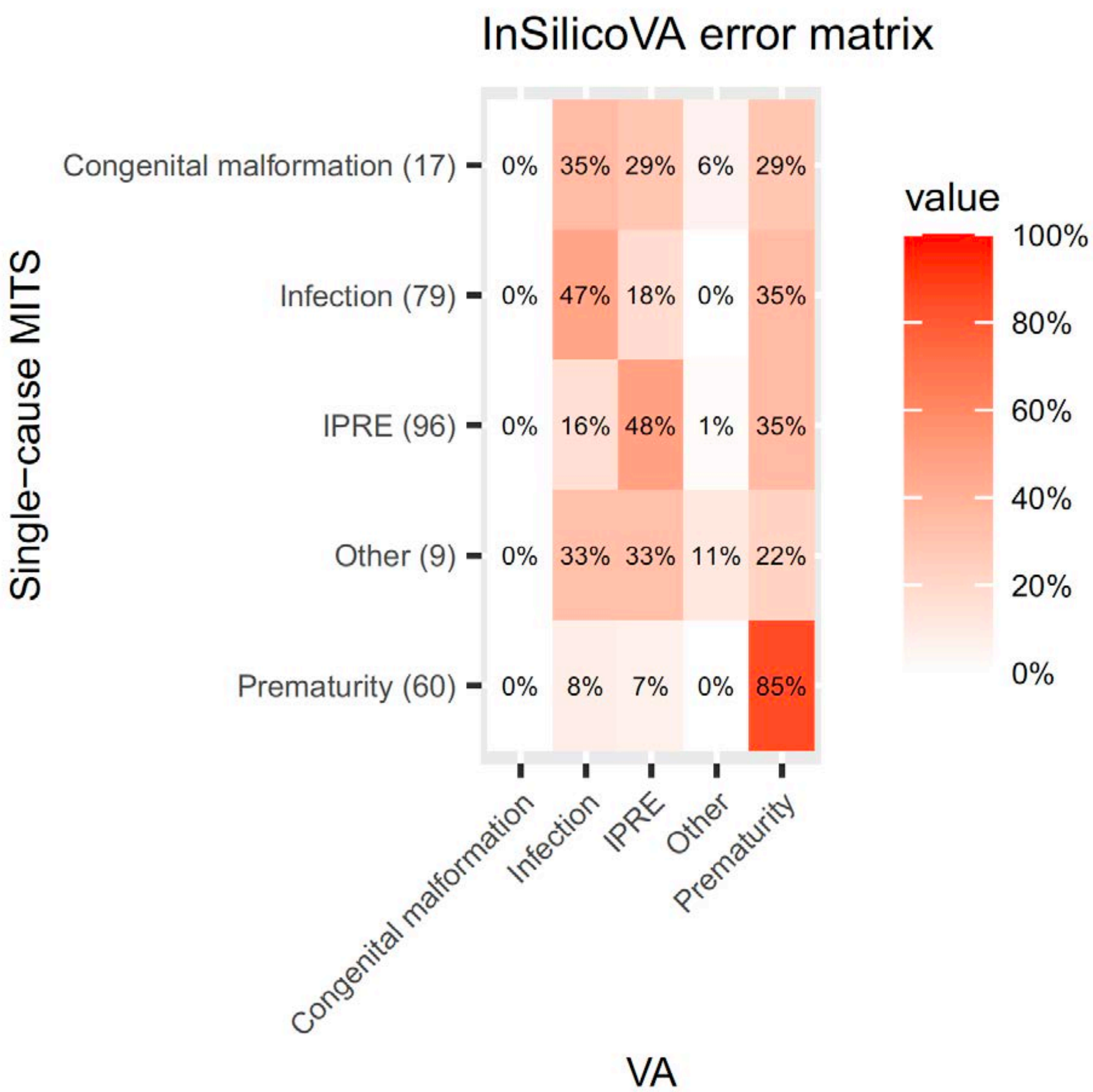
CHAMPS paired MITS-VA data

VA

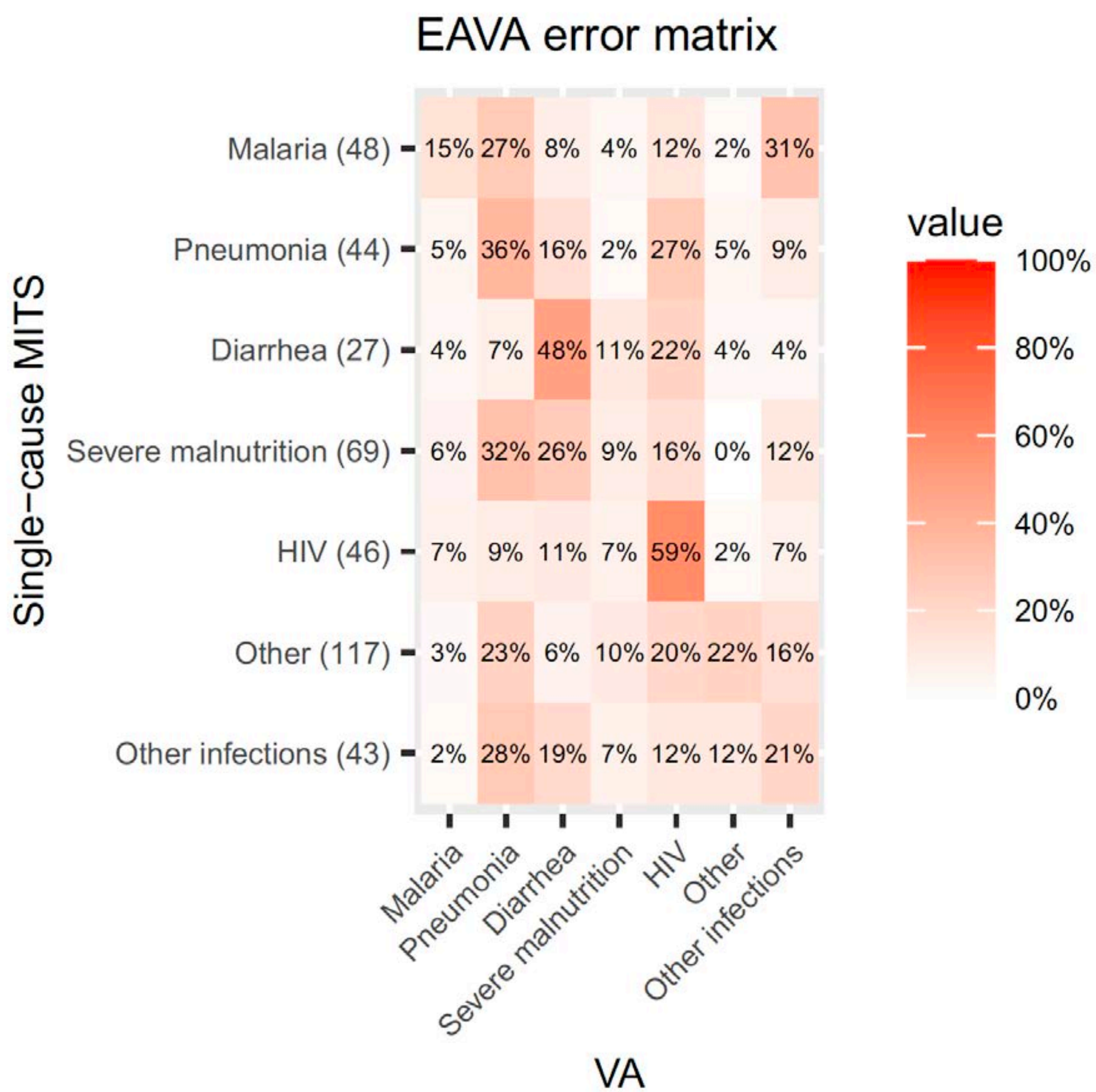
MITS		A	B	C
	A	100%	0%	0%
	B	0%	100%	0%
	C	50%	0%	50%

MITS-VA Misclassification rates matrix

Cause of death misclassification by VA



Misclassification rates of VA
for neonates in COMSA-Mozambique



Misclassification rates of VA
for under-5 children in COMSA-Mozambique

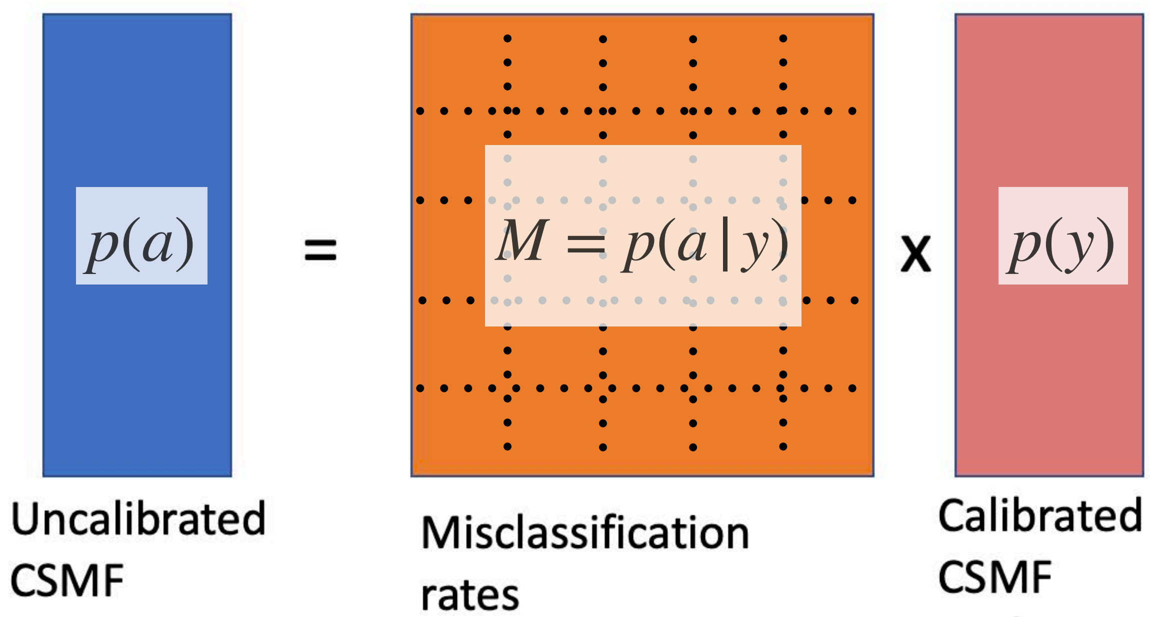
This paired data reveals that VA misclassifies COD in a large % of deaths

True for all age groups and choice of CCVA algorithm

VA calibration

Law of total probability: $$p(a) = \sum_y p(a \mid y) p(y)$$

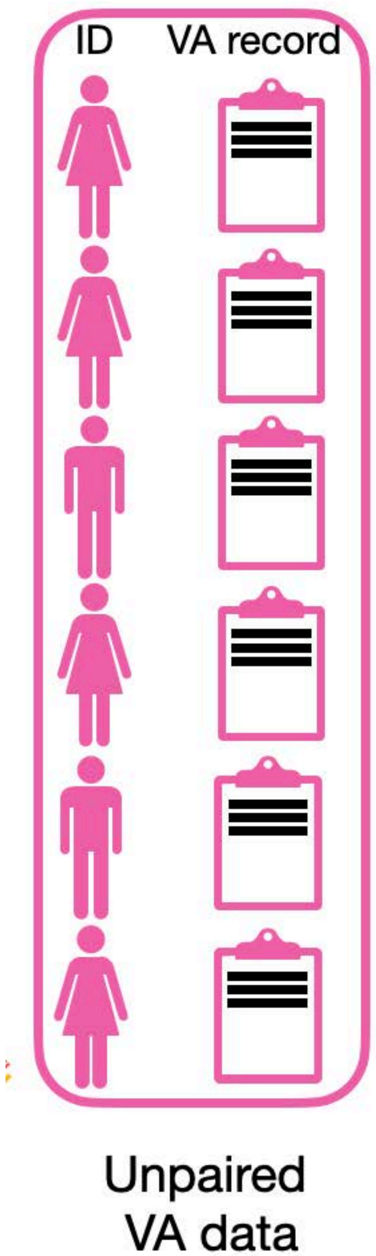
- *a* = VA COD, *y* = MITS COD, *M=[p(a=j|y=i)]* is the misclassification rate matrix

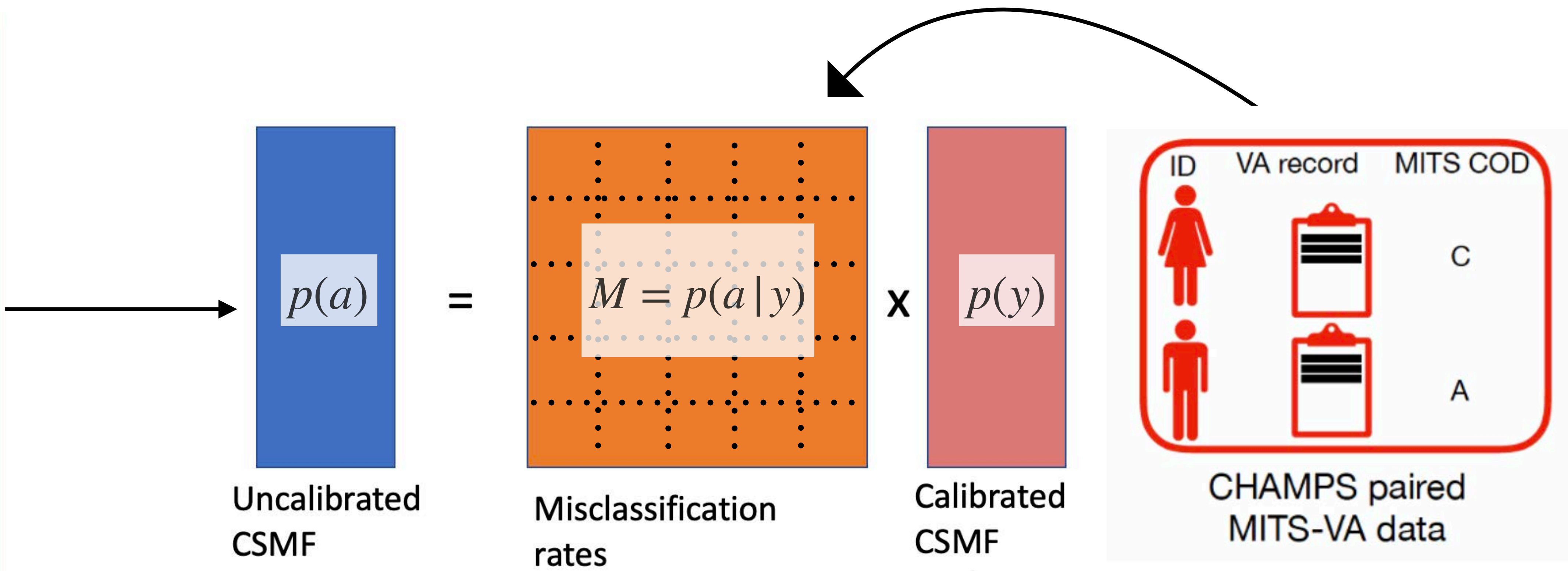


VA calibration

Law of total probability: $$p(a) = \sum_{y} p(a \mid y)p(y)$$

- *a* = VA COD, *y* = MITS COD, *M=[p(a=j|y=i)]* is the misclassification rate matrix



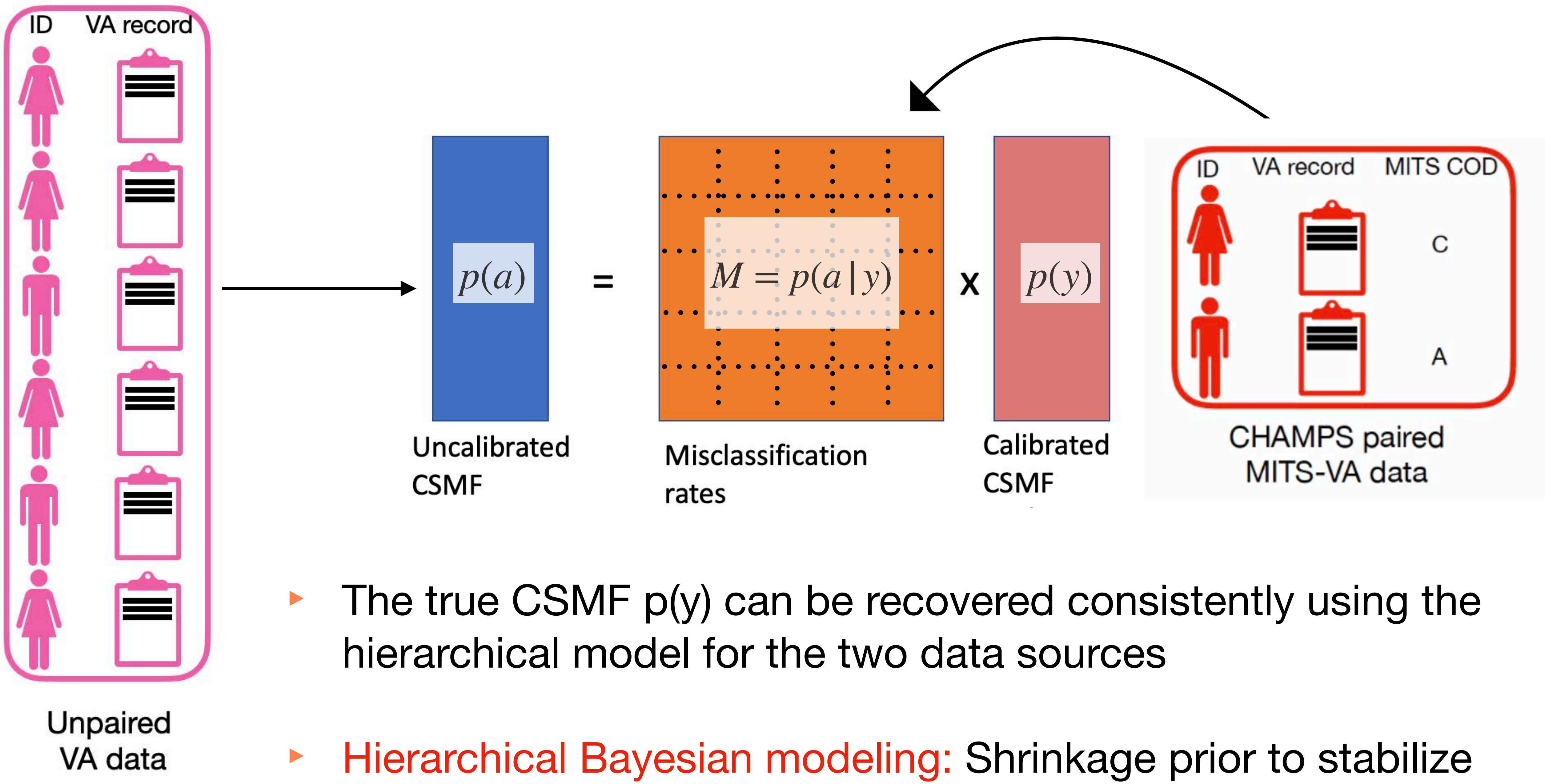


- Multinomial model for the unpaired COMSA VA data: estimates p(a)
- Multinomial model for the paired CHAMPS data: estimates the misclassification rates p(a|y)

VA calibration

Law of total probability: $p(a) = \sum_y p(a \mid y) p(y)$

- a = VA COD, y = MITS COD, $M=[p(a=j|y=i)]$ is the misclassification rate matrix



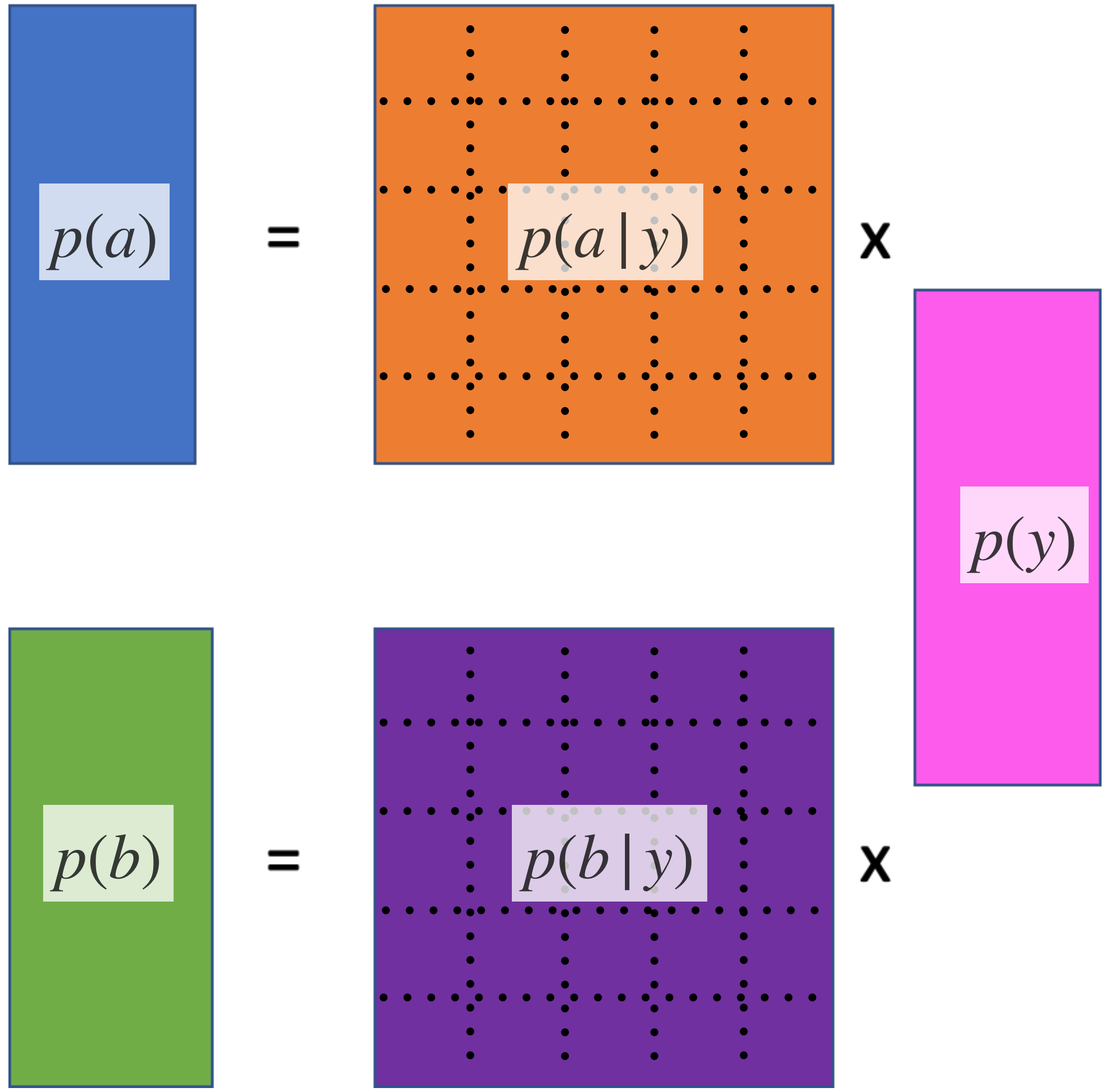
- The true CSMF p(y) can be recovered consistently using the hierarchical model for the two data sources
- Hierarchical Bayesian modeling: Shrinkage prior to stabilize estimation of the misclassification rates

Ensemble VA calibration

Many VA algorithms exist, each produces its own uncalibrated and calibrated CSMF

Ensemble calibration: Combines data from two (or more) VA algorithms into a single calibrated CSMF estimate.

Weights the algorithms in a data-driven way, favoring the more accurate ones.



Multiple VA-COD

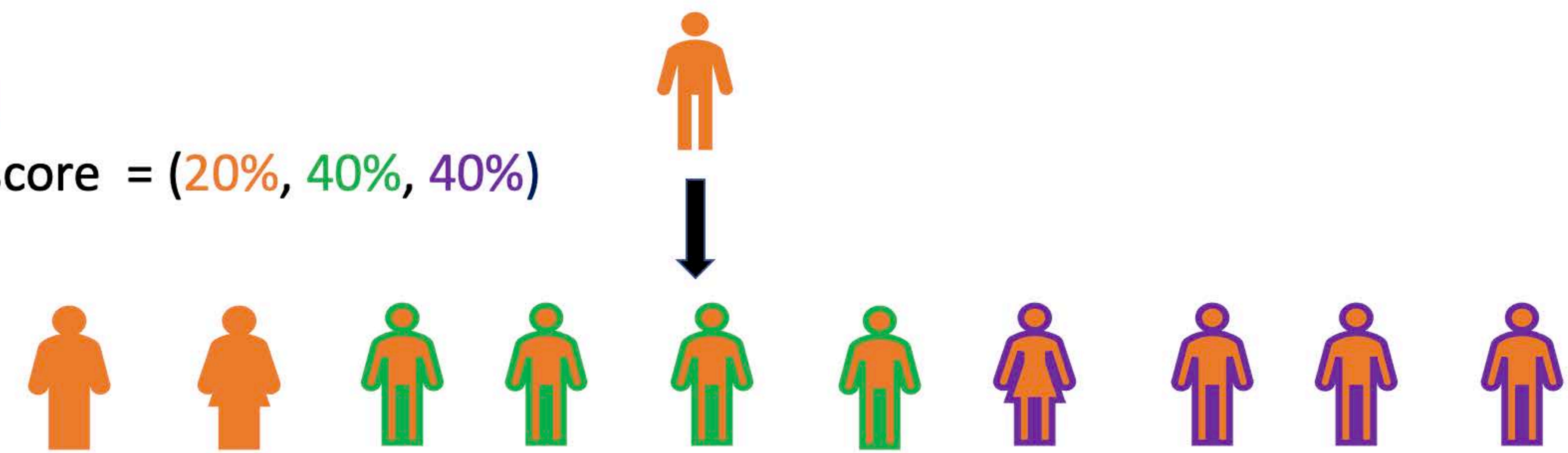
Many VA algorithms like InSilicoVA are probabilistic -- assigning scores to each cause being the underlying COD, i.e., compositional (or fractional) COD data

Create 10 copies of the VA record for each individual with causes of death assigned proportionately using the VA scores

Example:

True cause ■

VA predicted score = (20%, 40%, 40%)



All 10 copies will have the same MITS cause ■

Down-weight these copies by 1/10 in the calibration to keep the sample size same

Multiple MITS-COD using generalized Bayes

- MITS diagnosis for many cases can also suggest multiple cause of death
  - Underlying and Immediate causes

- So both VA- and MITS-COD are compositional data

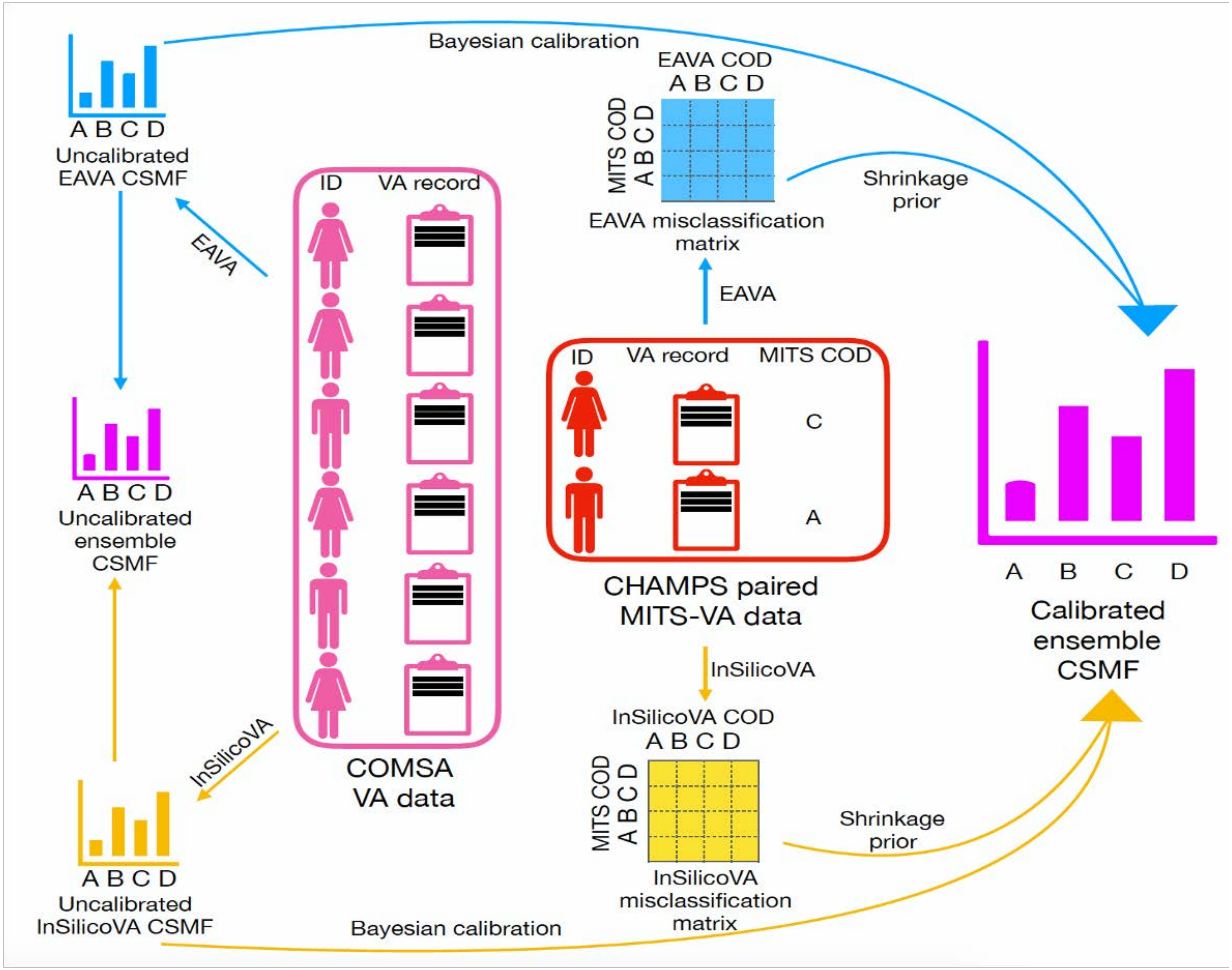
- Extend (weighted) multinomial likelihood to a Kullback-Liebler loss-function (pseudolikelihood) for composition-on-composition regression

- Generalized Bayes: Using any loss function $\ell(\theta \,|\, \text{data})$ and a priori $\pi(\theta)$

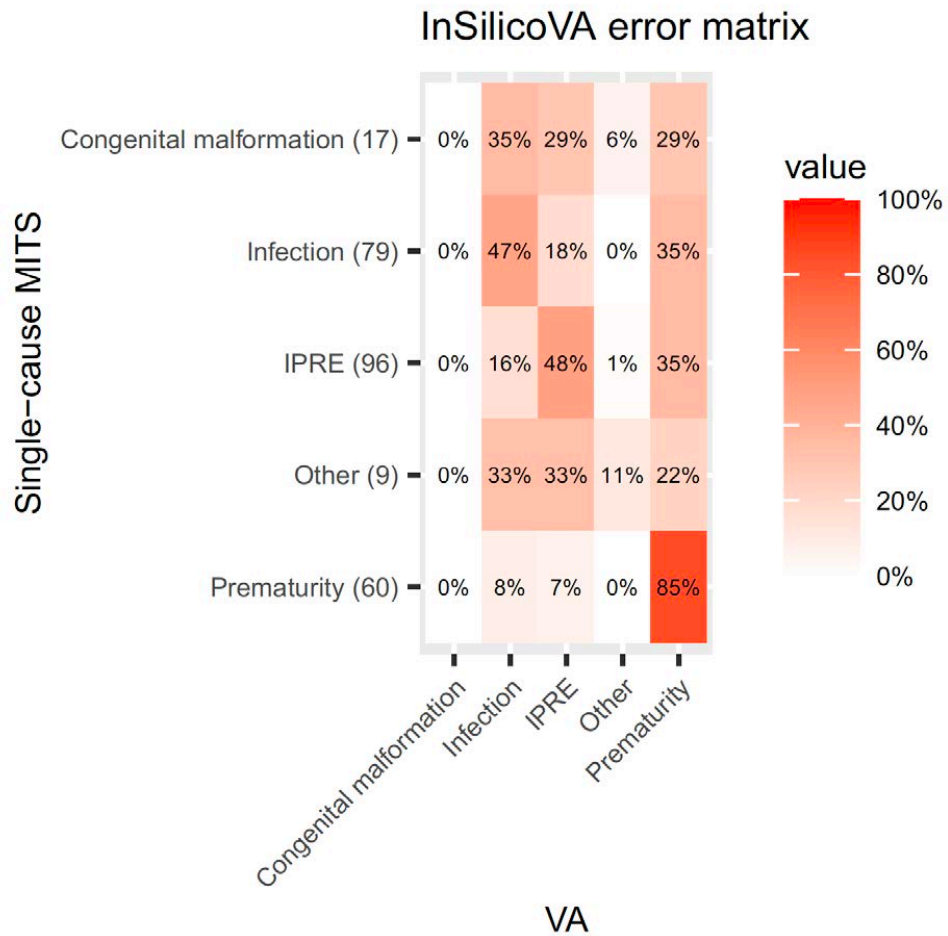
- Posterior: $\Pi(\theta \,|\, data) \propto \exp\left(-\ell(\theta \,|\, data)\right) \Pi(\theta)$

- Unifies categorical and compositional COD data types

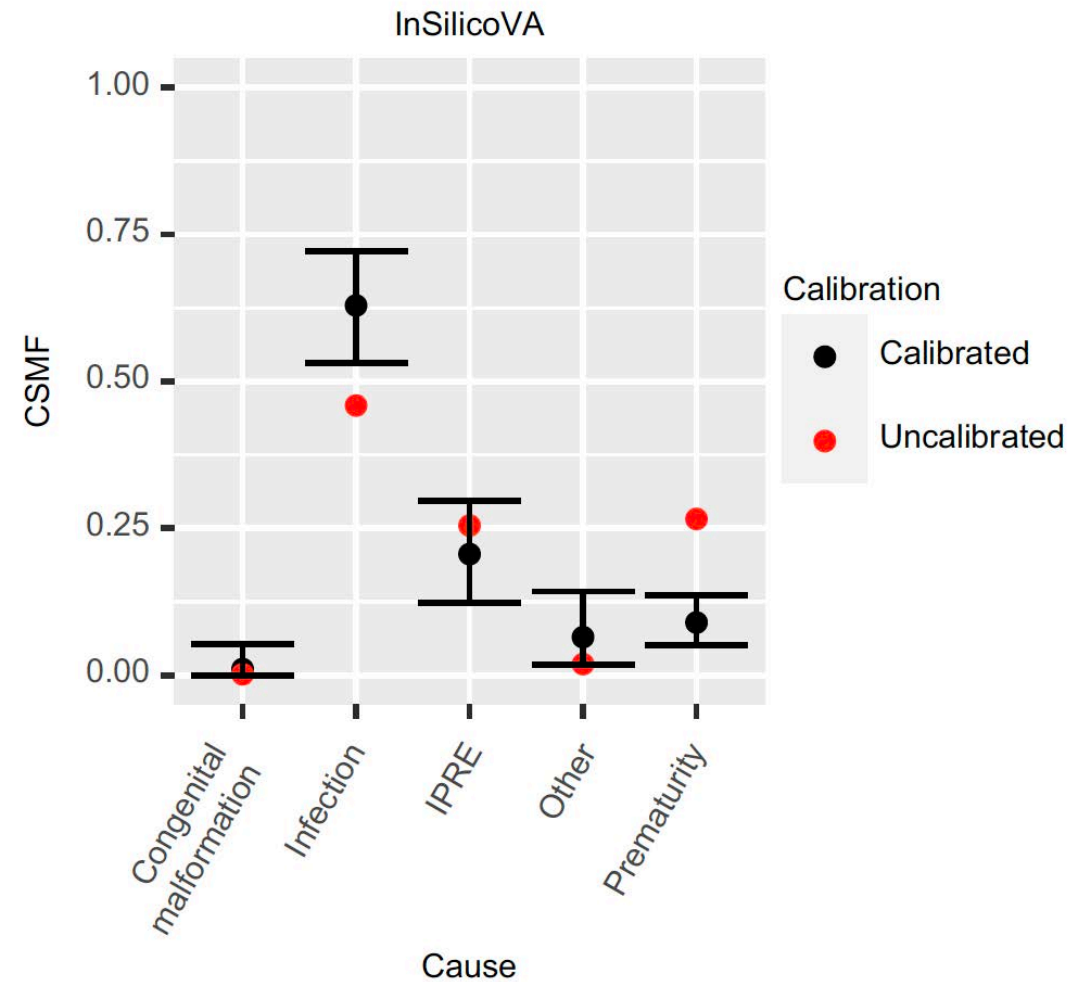
Overview of VA calibration pipeline for COMSA-Mozambique



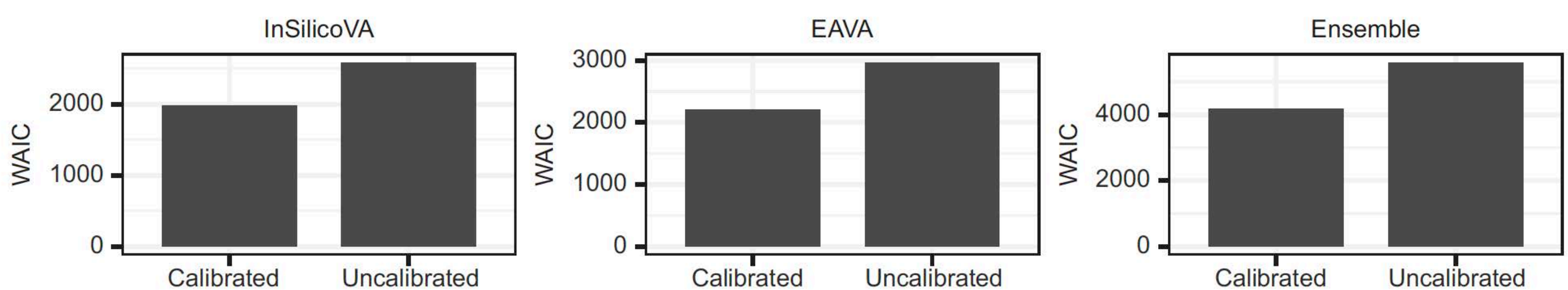
COMSA-Mozambique neonatal results



Misclassification rates of InSilicoVA for neonates in COMSA-Mozambique



Calibrated and uncalibrated CSMF



Model comparison of calibrated and uncalibrated CSMF using Widely Applicable Information Criterion WAIC

Ongoing work

- Transportability of MITS based VA-calibration for data from other countries
  - Account for cross-site heterogeneity in misclassification rates
- Global application of VA-calibration for improved child mortality estimates (CA-CODE, Pi: Liu)
- VA-calibration using summary data
- Sub-group specific VA-calibration (downscaling)
- Use cause-hierarchy to calibrate at a finer cause-resolution

Papers and software

- (Single-cause VA calibration) Datta, A., Fiksel, J., Amouzou, A., & Zeger, S. (2020). *Regularized Bayesian transfer learning for population-level etiological distributions.* Biostatistics, 1465–4644, 2020.
- (Multi-cause VA calibration) Fiksel, J., Datta, A., Amouzou, A., & Zeger, S. (2021). *Generalized Bayes quantification learning under dataset shift.* Journal of the American Statistical Association, 117(540), 2163-2181
- (Composition-on-composition regression) Fiksel, J., Zeger, S., & Datta, A. (2021). *A transformation-free linear regression for compositional outcomes and predictors.* Biometrics 78(3), 974–987, 2022.
- (Software) R-package *calibratedVA* for calibrating VA-based CSMF estimates using GBQL *https://github.com/jfiksel/CalibratedVA/*
- (Overview of VA-calibration) Fiksel J, et al. *Correcting for verbal autopsy misclassification bias in cause-specific mortality estimates* American Journal of Tropical Medicine and Hygiene, 105(5S), 66-77, 2023.
- *(Multi-cause VA-calibration for COMSA-Mozambique)* Gilbert B et al. *Multi-cause calibration of verbal autopsy-based cause specific mortality estimates of children and neonates in Mozambique* American Journal of Tropical Medicine and Hygiene, 105(5S), 78-89, 2023.
- *(COMSA-Mozambique main results)* Macicame I et al. *Countrywide Mortality Surveillance for Action in Mozambique: Results from a National Sample-Based Vital Statistics System for Mortality and Cause of Death* American Journal of Tropical Medicine and Hygiene, 105(5S), 5-16, 2023.

Thank you

abhidatta.com abhidatta@jhu.edu @datta_science