Using Post-Test Odds for Etiological Predictions of Infectious Diarrhea in Resource-limited Settings

Ben Brintz @brintzbe

Postdoc
Leung Lab @GlobalGutHealth

April 14, 2019

HEALTH
UNIVERSITY OF UTAH

Division of Infectious Diseases

Overview

1. Background
2. The Clinical Prediction Rule
   - Post-test Odds Formulation
   - Model/Test Building Process
3. Results
4. Discussion

For management of diarrhea in lower- middle- income countries (LMICs), a paucity of decision support tools leads to over-prescription of antibiotics

- In LMICs, etiological diagnosis is rarely made
  - cost constraints
  - availability of testing
- Treatment of diarrhea is commonly based on clinical suspicion
- Leads to toxicity, increased costs of care, and resistance

Recent studies have made etiological data available from large cohorts of children with diarrhea

- [Kotloff et al., 2013] - GEMS data published
- [Liu et al., 2016] - GEMS with qPCR
- [Platts-Mills et al., 2018] - MAL-ED with qPCR
- unpublished - VIDA (3 countries from GEMS) with qPCR

These studies provide a unique opportunity to derive etiological prediction rules

- GEMS is a prospective, case-control study in 7 countries (2007-2011)
- 9439 children with moderate-to-severe diarrhea enrolled and matched
- A fecal sample was taken from each child at enrollment to identify enteropathogens along with well characterized clinical information (>1000 variables)

We make two etiological predictions using multiple data sources to guide providers dealing with infectious diarrhea

- Viral-only etiology vs. Other known etiologies
  - Predicting this indicates the patient likely does not need antibiotics
- Any bacterial etiology vs. Other known etiologies
  - Predicting this indicates that antibiotics should be considered, and/or further testing may be warranted

Our goal is to build an electronic clinical decision support system (eCDSS) with multiple data sources appropriate for use in LMICs

- Easy to use interface for providers (parsimonious model)
- New or changing data sources
- Transient internet connection

We use the post-test odds formula to flexibly include various data sources into one test

- Data sources we considered: Clinical predictors, local aggregate of prior clinical information, local climate trends
- The model must be able to make predictions when not all data sources are available
- Post-test Odds Construction (with conditional independence assumption):

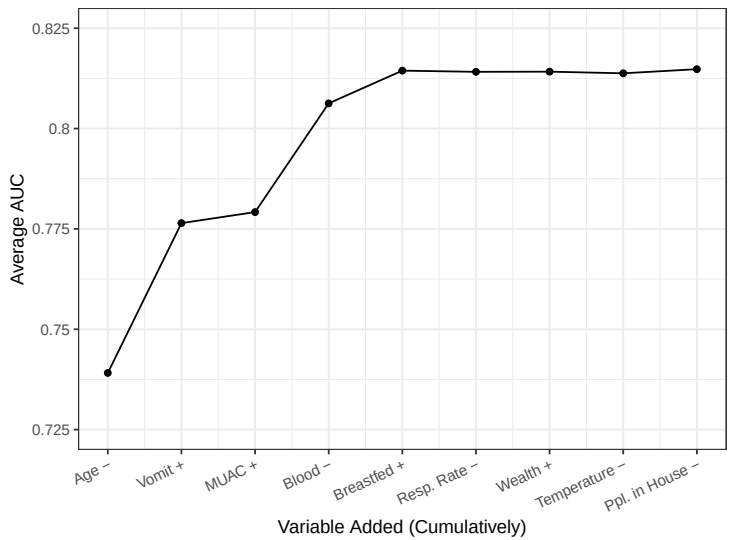
$$\begin{aligned}\frac{P(V=1|T_1=t_1,T_2=t_2,\cdots,T_k=t_k)}{P(V=0|T_1=t_1,T_2=t_2,\cdots,T_k=t_k)} &= \frac{P(V=1,T_1=t_1,T_2=t_2,\cdots,T_k=t_k)}{P(V=0,T_1=t_1,T_2=t_2,\cdots,T_k=t_k)} \\ &= \frac{P(T_1=t_1,T_2=t_2,\cdots,T_k=t_k|V=1)\cdot P(V=1)}{P(T_1=t_1,T_2=t_2,\cdots,T_k=t_k|V=0)\cdot P(V=0)} \\ &= \frac{P(V=1)}{P(V=0)}\cdot\prod_{j=1}^{k}\frac{P(T_j=t_j|V=1)}{P(T_j=t_j|V=0)}\end{aligned}$$

- Along with pre-test odds, any test can be included or excluded in the product as a conditional likelihood ratio without re-training

T1: GEMs variables are useful for developing predictive models based on clinical information to be applied to future diarrhea patients

- Cleaned data by removing variables
  - related to controls
  - not available at the time of presentation
  - redundant variables
  - 160 left
- Used variable importance measure (reduction in variance) using Random Forest regression
- Fit models of various parameter sizes with current patient predictors (logistic regression and RF regression)

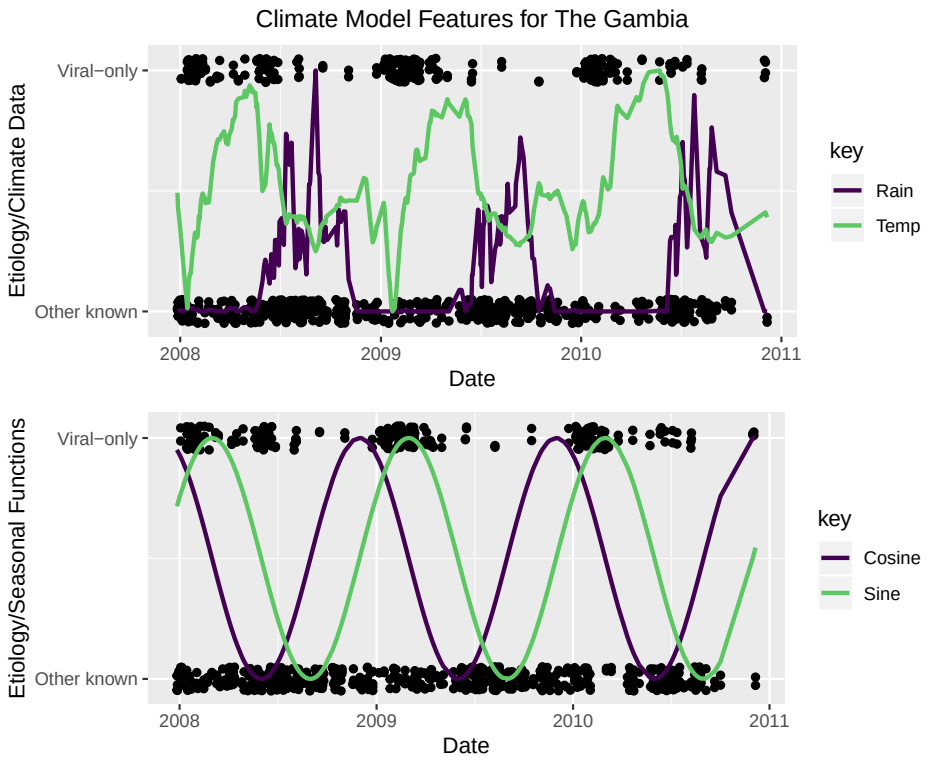
Variable Name	Avg. Variance Dec.
Age	46.10
Vomit	22.04
BMI	21.32
Blood	20.84
Breastfed	19.04
MUAC	18.63
Resp. Rate	15.33
Wealth	14.66
Temperature	13.57
Ppl. of House	10.95



T2: Seasonal patterns in etiology are accounted for with climate data

- Obtain temperature and rain data from local NOAA weather station (uploaded daily)
- Create two-week moving average to represent seasonality
- Use seasonal sine and cosine curves

  $sin(\frac{2\pi t}{365.25})$ and $cos(\frac{2\pi t}{365.25})$
- Logistic regression



T3: An aggregate of prior clinical responses can represent a clinician's intuition

- Clinicians say that if they observe an influx of patients with a certain clinical presentation (such as younger age, with vomiting - i.e. viral), might expect the next patient to have similar etiology

- We use a weighted average of recent clinical information (two weeks) collected in GEMS to represent that intuition

$$\overline{W}_{age(d)} = \frac{\overline{age}_{d-1} * w_1 + \overline{age}_{d-2} * w_2 + \cdots \overline{age}_{d-n} * w_n}{w_1 + w_2 + \cdots w_n}$$ where w_i are from the Wendland family of covariance functions

$$\overline{W}_{Blood(d)} = \frac{\overline{W}_{T_{blood}}}{\overline{W}_T}$$

- Fit a logistic regression model with weighted average variables and seasonal sine and cosine

For each data source, conditional likelihood ratios $\frac{P(T_j=t_j|V=1)}{P(T_j=t_j|V=0)}$ are developed from training data using Gaussian kernel density estimates

- Example of current patient data test (logistic regression)

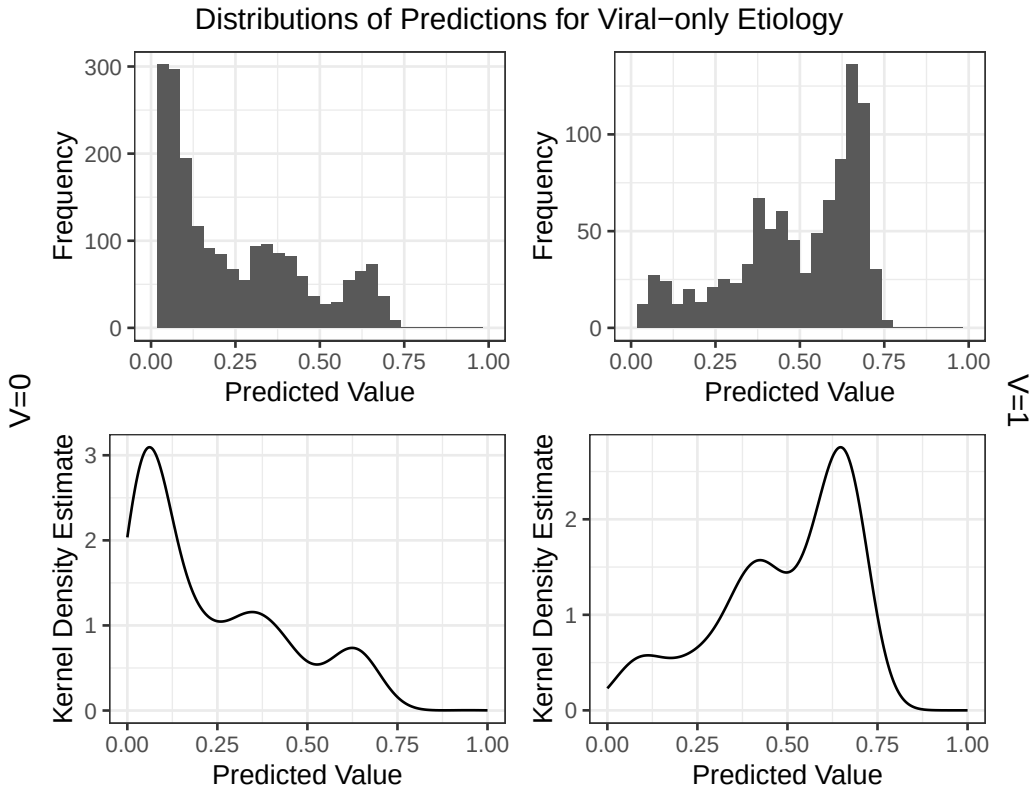


Figure: Histograms and estimated kernel densities of predicted values obtained from logistic regression on GEMS patient training data. The left graphs represent other known etiologies and the right graphs represent viral etiologies.

The post-test odds conditional independence assumption can be overcome using a joint density

- Example of 2D Gaussian kernel density estimates

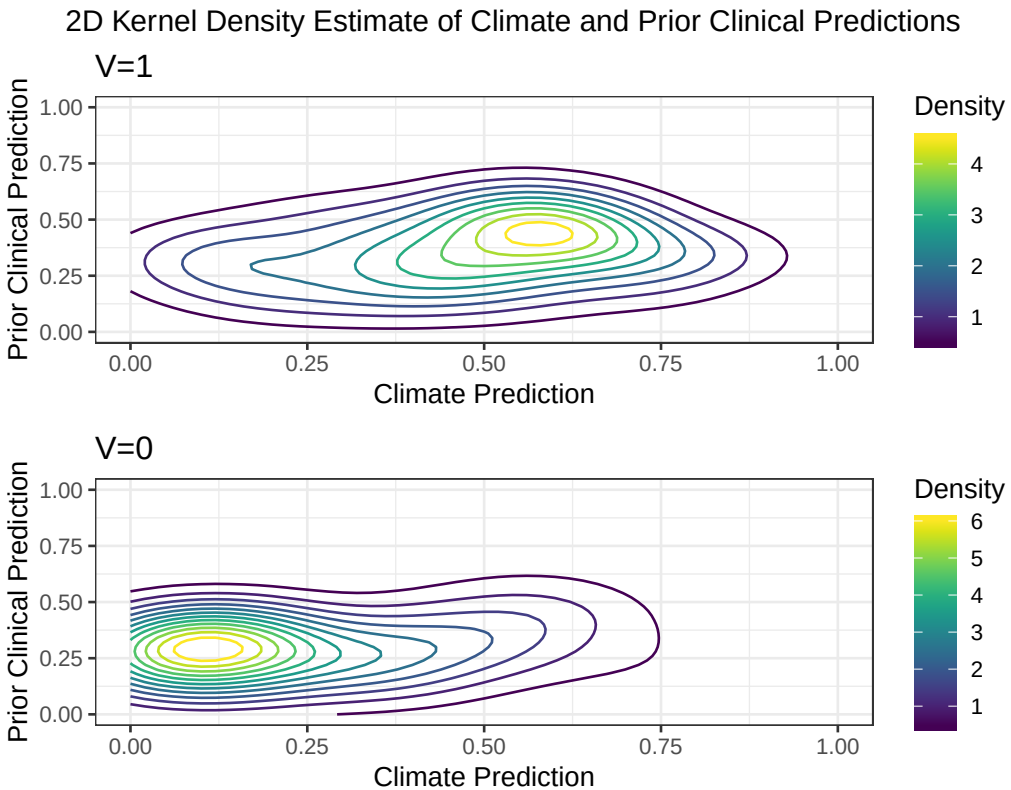


Figure: Contour plots of 2-dimensional kernel densities of predicted values obtained from logistic regression on GEMS prior patient data and climate data. The top graph represents viral etiologies and the bottom graph represents other known etiologies.

We used 5-fold cross-validation on the Post-Test Odds model to obtain generalizable results (200 iterations)

- Results from logistic regression with only clinical predictors

Current Patient Only/LR only	
All countries in GEMs	VIDA countries in GEMS
.8113	.7706

- All results include pre-test odds (training set odds of viral only) and current clinical test

Climate	Prior Clinical	Joint Climate/Prior Clinical	Avg. AUC
All countries in GEMS			
✓			0.8512
	✓		0.8502
✓	✓		0.8362
		✓	0.8470
VIDA countries in GEMS			
✓			0.8199
	✓		0.8171
✓	✓		0.8079
		✓	0.8189

We externally validated with the VIDA data set

- Results from logistic regression with only clinical predictors

Current Patient Only/LR only
0.7238

- Tests trained on entire GEMS data set and tested on entire VIDA data set with known etiologies

Climate	Prior Clinical	Joint Climate Prior Clinical	Avg. AUC
✓			0.8035
	✓		.7947
✓	✓		0.8096
		✓	0.7984

Discussion

- The post-test odds methods performs well on externally validated data set
- Flexible enough for a transient internet connection or adding new tests
- Kernel density estimates for estimating likelihood ratio allows for uncertainty in predictions
- Joint density addresses conditional independence issue but is likely not need in practice
- Including climate and prior clinical trends allows for improved estimates

Future Directions

- Bayesian Networks for conditional tests and missing data
- Real-time updating on eCDSS for future unknown etiologies

Acknowledgments

- University of Utah, Salt Lake City
  - Joel Howard
  - Benjamin Haaland
  - Tom Greene
  - Andrew T. Pavia
  - Daniel T. Leung
- Institute of Disease Modeling, Seattle
  - Dennis L. Chao
  - Joshua L. Proctor
- University of Virginia, Charlottesville
  - James A. Platts-Mills
- University of Maryland, Baltimore
  - Karen L. Kotloff
- Funding
  - NIH R01 AI135114
  - BMGF grant OPP1198876

Questions?

Kotloff, K. L., Nataro, J. P., and Blackwelder, W. C. (2013).

Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the global enteric multicenter study, gems): a prospective, case-control study.

The Lancet, 382(9888):209–222.

Liu, J., Platts-Mills, J. A., Juma, J., Kabir, F., Nkeze, J., Okoi, C., Operario, D. J., Uddin, J., Ahmed, S., Alonso, P. L., et al. (2016).

Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the gems case-control study.

The Lancet, 388(10051):1291–1301.

Platts-Mills, J. A., Liu, J., Rogawski, E. T., Kabir, F., Lertsethtakarn, P., Siguas, M., Khan, S. S., Praharaj, I., Murei, A., Nshama, R., et al. (2018).

Use of quantitative molecular diagnostic methods to assess the aetiology, burden, and clinical characteristics of diarrhoea in children in low-resource settings: a reanalysis of the mal-ed cohort study.

The Lancet Global Health, 6(12):e1309–e1318.