Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Jessica Randall Date

Seroconversion Rate for Post-Validation Surveillance of Trachoma

By

Jessica M. Randall
MPH

Global Environmental Health

_____

Paul Weiss, MS

Committee Chair

_____

Matthew Freeman, PhD

Committee Member

_____

Diana Martin, PhD

Committee Member

Seroconversion Rate for Post-Validation Surveillance of Trachoma

By

Jessica M. Randall

B.S.
George Mason University
2014

Thesis Committee Chairs:

Paul Weiss, M.S.

An abstract of a thesis submitted to the Faculty of the Rollins School of Public Health of Emory University in partial fulfillment of the requirements for the degree of Master of Public Health in Global Environmental Health.
2019

Abstract

Seroconversion Rate for Post-Validation Surveillance of Trachoma

By Jessica Randall

Purpose*:* The WHO is considering the utility of serological surveillance for trachoma. If comparable to clinical diagnoses, this type of surveillance would be a cost-effective and sustainable solution. Our serological survey from two districts in The Republic of Togo, a country now known to have reached the threshold for elimination of trachoma as a public health problem, supports the use of a seroconversion rate as an estimate of force of infection in non-endemic areas.

Methods: We conducted serological surveillance on 2915 participants 1-9 years old in two districts purposefully selected as the most likely to have trachoma if it would be in the country at all. Each participant had a finger-prick blood sample collected onto filter paper. We fit a logistic regression model to estimate seroprevalence. Participants had blood collected to measure antibody responses to the *C. trachomatis* (CT) antigens Pgp3 and Ct694 by multiplex bead-based immunoassay (MBA), the lab-based dipstick Pgp3 lateral flow assay (LFA), and the field based cassette Pgp3 lateral flow assay.We calculated a seroconversion rate for each antigen by fitting a reversible serocatalytic model which assumes zero seroreversion. We report median fluorescence intensity-background (MFI-BG), seroprevalence, age-specific seroprevalence and seroconversion rate.

Results: Out of 2915 samples, 96% were negative by the MBA and LFA. Age-specific seroprevalence was estimated using a logistic regression model. Mean SCRs for each antigen and assay method were estimated by taking the mean of each age-specific seroconversion rate.

Conclusions: Our results support those of a parent study that found that the clinical diagnosis of trachoma, trachomatous inflammation- follicular (TF) falls below the 5% WHO threshold of elimination of trachoma as a public health problem. Our calculation of a seroconversion rate, an estimation of the force of infection, indicated no active transmission.

Seroconversion Rate for Post-Validation Surveillance of Trachoma

By

Jessica M. Randall

B.S.
George Mason University
2014

Thesis Committee Chair:

Paul Weiss, M.S.

A thesis submitted to the Faculty of the Rollins School of Public Health of Emory University in partial fulfillment of the requirements for the degree of Master of Public Health in Global Environmental Health.
2019

Acknowledgements

I would first like to thank Dr. Diana Martin, Mr. Paul Weiss, and Dr. Matthew Freeman for their feedback in the development of this manuscript and their guidance in placing the results in the larger context of the epidemiology of trachoma. Next, I would also like to thank Sarah Gwyn, Dr. Rachel Bronzan, the field teams of Health and Development International Inc., and the participants in the study. Finally, I would like to thank Alex Edwards, Ryan Wiegand, Dr. Lance Waller, Dr. Glen Satten, and Dr. David Nicholson

Table of Contents

Introduction

Approximately 2.2 million people are visual impaired or blind due to trachoma, the leading infectious cause of blindness worldwide (World Health Organization, 2016a). Repeated ocular infection with the bacterium *Chlamydia trachomatis* can lead to scarification of the inner eyelid, causing the lashes to turn in and scratch the eye. The World Health Organization (WHO) Alliance for the Global Elimination of Trachoma by 2020 (GET2020) endorses the SAFE strategy; Surgery to correct in-turned eyelashes, mass distribution of Antibiotics in affected regions, and behavior change promotion to improve Facial cleanliness and Environmental improvements The latter three interventions – A, F, and E – aim to reduce the population prevalence of the sign trachomatous inflammation—follicular (TF) to ≤5% in children aged 1–9 years (World Health Organization, 2016b)

A, F, and E interventions are initiated based on the outcomes of population-based prevalence surveys to determine the frequency of TF in an evaluation unit corresponding to a population size of 100,000–250,000 individuals (Solomon et al., 2015). If warranted, annual mass drug administration (MDA) of azithromycin is carried out for a set number of years based on the starting TF prevalence. Impact surveys are then conducted to determine if MDA can be stopped. At least two years after MDA stops, surveillance surveys are undertaken to determine if TF ≤ 5% is sustained. Once all previously trachoma-endemic areas have achieved and sustained TF ≤5% in 1-9-year-olds, and have met morbidity targets, the country can apply to the World Health Organization for validation of elimination of trachoma as a public health problem.

There is currently no guidance for post-validation surveillance, but serological surveillance using dried blood spots (DBS) to estimate antibody prevalence is one potential approach being evaluated for this purpose. Serologic surveillance would allow testing of samples collected for other purposes would eliminate the cost of a trachoma-specific survey for surveillance, and further reduce costs and resources required by integrating trachoma surveillance into other health program activities ((S. E. Gwyn et al., 2018). We have previously evaluated antibody responses to the immunodominant *C. trachomatis* antigens Pgp3 and CT694 using a multiplex bead assay (MBA) and the Pgp3 using a lateral flow assay (LFA) (E. Brook Goodhew et al., 2012; West et al., 2016). The MBA allows for integration of serologic surveillance of a number of diseases in a single well of a 96 well specimen plate using less than a drop of blood and provides semi-quantitative data reflective of the intensity of the antibody response. The LFA is easy to use and does not require extensive instrumentation, but provides only dichotomous positive or negative results (E. B. Goodhew et al., 2012). Using results from these assays along with demographic data, we can estimate the age-specific seroprevalence of *C. trachomatis* antibody response that previous work supports as a proxy measure of transmission (Martin et al., 2015).

In 2017, Health & Development International (HDI) undertook population-based prevalence surveys in 7 provinces in the Republic of Togo to determine if interventions for trachoma were required. In an as of yet unpublished paper, in all settings, TF was <5%. As part of this survey, blood was collected by field teams, who created dried blood spots (DBS) and field-tested the Pgp3 LFA(Sarah Gwyn et al., 2016). DBS were sent to

the U.S. Centers for Disease Control and Prevention to be analyzed by MBA for antibody responses to Pgp3 and CT694 and for antibodies to Pgp3 on laboratory-based Pgp3 dipstick LFA in order to assess the utility of serological tools for post-validation surveillance (S. Gwyn, Mkocha, Randall, Kasubi, & Martin, 2019).

Methods

Ethics Statement: Institutional Review Board (IRB) approvals were obtained for collections of DBS by the Ministry of Health and Social Protection, National Program for the Control against Blindness, Togo (013/2016/CBRS).). CDC researchers were non-engaged in the study and did not interact with study participants or have access to any identifying information. CDC researchers were non-engaged in the study and did not interact with study participants or have access to any identifying information. Parental consent was obtained for all participants. Samples were collected and used with permission by Human Research International.

DBS collection: From the two districts selected for DBS collection the Health & Development International (HDI) team collected a finger-prick blood sample onto filter paper (Trop-Bio, Townsville, Australia) using a sterile single-use lancet (BD Microtainer, Dublin, Ireland). Each filter paper had six extensions, calibrated to absorb 10 µL of blood each. Samples were air-dried for approximately five hours and then placed in individual Whirl-Pak plastic bags (Nasco, Modesto, California) which were stored with desiccant sachets (Whatman, Little Chalfont, UK) at −20 °C. All samples were shipped to CDC for testing (Migchelsen et al., 2017).

Field LFA Testing: After representatives of the Health Research International collected finger-prick blood onto DBS, additional blood was collected into a 10 µL micropipette (CAT#) and placed on the sample port of the Pgp3 LFA (Sarah Gwyn et al., 2016; S. Gwyn et al., 2019). Three drops of buffer was added and cassettes were evaluated after 30 min for the presence or absence of a test line and control line. Samples were marked as positive, negative, or invalid (if no control line).

Multiplex Bead Assay: One DBS extension containing 10 µL of blood was eluted overnight at 4°C into 1.6 mL of buffer solution. DBS eluates were screened in duplicate with Pgp3-coupled beads on the multiplex bead assay (MBA) as previously described (Goodhew, 2012). Data were reported as median fluorescence intensity (MFI) with the background from blank control wells subtracted out (MFI-BG) for each antigen for each sample. Cutoffs for antibody positivity were determined by receiver operator characteristic curves using previously characterized samples. The cutoff in MFI-BG for positivity by Pgp3 was 1647 and the cutoff for positivity by CT694 was 347.

Pgp3 dipstick LFA. One DBS extension was eluted in 60 µL of LFA buffer for 4-24 hours at 4°C in a flat-bottom 96-well plate (USA Scientific, Ocala, Florida, USA). The conjugate master mix was prepared at a volume proportional to the number of wells to test and 20 µL of conjugate mixture was added to each well no more than 4 hours prior to testing. LFA-dipsticks were inserted into each well until the DBS eluate-conjugate mixture was absorbed, approximately 15-20 minutes. Forty µL of LFA buffer was added

to each well to “chase” the background caused by hemolysed red blood cells. LFAs were read as positive, negative or invalid after the membrane was completely cleared (5-10 minutes). Examples of positive, negative, and invalid tests have been described elsewhere (S. Gwyn et al., 2019).

Statistical analyses:

In this analysis the software packages Microsoft Excel 2010, SAS 9.4, and Python 3.7 were used for descriptive statistics, data visualization, and logistic regression. The 95% confidence intervals associated with the seroprevalence were calculated using the Adjusted Wald or Wilson method and are defined as the number of positive tests divided by the number of valid samples tested (MeasuringU, 2018).

Participants: We merged datasets containing demographic and field results from Health and Development International and lab results from CDC by ID number for 2986 unique ID numbers. We then removed those ID numbers where we found duplicates, invalid tests, those without demographic information and 1 sample ID which was a duplicate run on the MBA, invalid on the dipstick LFA, and did not have any associated demographic information (n=67). Those remaining ID no. with complete data (n=2915) were included in the Lab set for analysis. Approximately 2/3 of the samples which were run in the lab were also run in on the LFA in the field. To analyze these field results we started with the lab set of 2915 and excluded those ID nos. which were not run in the field (n=1002) and those sample IDs which were excluded from the lab dataset due to duplicate sample IDs, missing demographic data, or invalid test results from their lab tests (n=48). The resulting Field set contained 1914 complete entries.

The primary outcome of interest was trachoma seroprevalence, represented by a positive or negative result by the Pgp3 and CT694 antigens on the MBA and the Pgp3 antigen of the dipstick and cassette LFAs.

Positive results were coded and 1 and negative results were coded as 0. A MFI reading ≥1647 by Pgp3 was coded as positive and a MFI reading by CT694 ≥ 357 was coded as positive. All other MFI results were coded as negative. MFI results from the Lab dataset were log transformed to approximate normality.

We fit a descriptive logistic regression model to estimate seroprevalence by each antigen and assay method. From these estimated probabilities we calculated a seroconversion rate by fitting a reversible serocatalytic model which assumes zero seroreversion. We report MFI-BG, seroprevalence, age-specific seroprevalence and SCR. Tables containing additional analyses as well as the SAS and Python code generated for the project are available in the Appendix.

Results

Descriptive statistics were run on age, sex, province, and positivity results by each antigen from both the Lab and Field datasets. Age was used in a logistic regression model to estimate the proportion of positive samples as determined by MBA (Pgp3), MBA (CT694), and the Pgp3 dipstick LFA. If $\hat{Y}$ is the estimated seroprevalence of trachoma let

$$\text{Estimated Seroprevalence} = \hat{B}0 + \hat{B}1 * (\text{age in years}) + \text{standard error}$$

Age-specific results from this logistic regression model were fit to a serocatalytic model where λ is the estimated seroconversion rate with an assumed seroreversion rate of 0.

Estimated Seroprevalence= (ln (Estimated Trachoma Seroprevalence-1))/Age in Years

Of 2915 enrolled individuals in the Lab set, 1504 (50.2%) were from province number 40001, 1411 (48.4) were from province number 40002. Of the total, 1446 individuals (49.6%) were female. Of 1914 enrolled individuals in the Field set, 967 (50.5%) were from province 40001, 947 (49.5) were from province 40002. Of the total, 966 individuals (50.5%) were female.

MFI-BG: Raw MFI values ranged from -4 to 31469 so in order to more clearly visualize every point, data were log transformed to range from 0 to 4.50. Out 2915 samples, 96.5% (n=2,812) fell below the cut-off values, 2.5 for CT 694 and 3.2 for Pgp3.

Seroprevalence using complete lab dataset: Of 2915 samples 103 (3.5%, 95% CI 2.9-4.3) tested positive for the presence of anti-Pgp3 and anti-CT694 antibodies on the MBA, and 119 (4.1%, 95% CI 3.4-4.9) had a positive dipstick LFA result.

Seroprevalence comparisons with field dataset: Of 1914 samples, 69 (3.6%, 95% CI 2.9-4.5) tested positive for the presence of anti-Pgp3 antibodies on the MBA, 68 (3.6%, 95% CI 2.8-4.5) tested positive for anti-CT694 antibodies by MBA, 81 (4.2%, 95% CI 3.4-5.2) had a positive dipstick LFA result, and 100 (5.2%, 95 %CI 4.4-6.3) had a positive cassette LFA result.

Age-specific seroprevalence and Seroconversion Rate Estimates: Age-specific seroprevalence was estimated using a descriptive logistic regression model assuming 100% sensitivity. Code and datasets used for these estimations are available in supplementary documentation. The resulting estimated seroprevalence estimates were fit to a serocatalytic model which assumed no seroreversion, described elsewhere (Drakeley et al., 2005). Mean SCRs for each antigen and assay method were estimated by taking the mean of each age-specific seroconversion rate. These means are displayed below overlaid on graphs of the estimated age-specific seroprevalence. Rates across all ages were approximately zero.

Discussion

Previous work suggests the utility of the SCR as an approximation of force of infection of trachoma (Drakeley et al., 2005). SCRs of <0.015 per year have also been shown to correspond to TF <5% (Pinsent et al., 2018). Our data suggest that these two provinces have <5% seropositivity by both the MBA and LFA. Our work supports previous findings in the use of SCR as an approximation of TF since our estimated SCRs were between 0.0070-0.0099 and baseline mapping revealed that those districts in which we collected DBS were already below the elimination threshold of 5% TF. We believe our estimates may have been improved by controlling for interaction effects between age and seropositivity by other antigens used. When estimating seroprevalence based on reactivity to Pgp3, the model did not control for interactions between positivity by CT694 and age and the reverse was true for the models estimating seroprevalence by CT694 and the Dipstick LFA. Even with potential improvement in precision in studies where more

than one antigen is used, our estimates support the use of the SCR to estimate force of infection and approximate TF in instances where only one antigen is used.

We also saw an increase in estimated seroprevalence and the resulting SCR with increases in age. However, since the SCR approximates zero it is unclear if this pattern would be observed regardless of transmission intensity. Future research could explore whether this increase in reactivity with age would also appear in a population that has never been exposed to C.*trachomatis* bacteria.

Looking at the MFI-BG heat map, the majority of samples' results fell well below the cut-offs for positivity. However, it is hard to say with certainty whether or not any of the 3.5% of samples that were counted as positive by the MBA were truly positive. In a similar study among 1-9 year olds in Bolivia and the United States, the sensitivity of the MBA was 93.2% (CI 88.3-98.1) and the specificity ranged from 96.1% (CI 91.8-98.2) to 99.4% (CI 98.2-100) (Sarah Gwyn et al., 2017). Our own results fall within these ranges so any positive results must be interpreted with caution. A study looking at the LFA when used with whole blood specimens reported sensitivity to be 89% (CI 82.0-96.0) and specificity to be 96% (CI 93.0-99.0) (Wiegand et al., 2018). Future work in similar settings should be cautious in their interpretations of results as truly representative figures of exact seroprevalence.

A subset of 1914 unique samples were run on the cassette LFA in Togo to assess the performance of the assay under different conditions. This cassette LFA showed

comparable results to the samples tested at CDC on the MBA and dipstick LFA though field results did tend to be higher than lab based test results. LFA field deployment poses unique challenges. For example, research teams may be moving from house to house quickly and reading the tests at varying time intervals. Even with these challenges our results suggest that the LFA cassette could provide field teams with results somewhat comparable to those found by the MBA or dipstick LFA.

Conclusion and Recommendations

Our results support those of a parent study confirming that Togo has indeed met the WHO threshold for validation of elimination of trachoma as a public health program and the use of the SCR as a metric for serosurveillance in post-validation settings. The field-based cassette LFA also provides comparable results to the lab-based MBA and dipstick LFA. In looking at future post-validation settings it will be important to be cognizant of the limitations of interpretations of results which may fall within the sensitivities and specificities of the assays used.

References

Drakeley, C. J., Corran, P. H., Coleman, P. G., Tongren, J. E., Mcdonald, S. L. R., Carneiro, I., . . . Riley, E. M. (2005). Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proceedings of the National Academy of Sciences*.

Goodhew, E. B., Priest, J. W., Moss, D. M., Zhong, G., Munoz, B., Mkocha, H., . . . Lammie, P. J. (2012). CT694 and pgp3 as Serological Tools for Monitoring Trachoma Programs. *PLoS Neglected Tropical Diseases*. doi:10.1371/journal.pntd.0001873

Goodhew, E. B., Priest, J. W., Moss, D. M., Zhong, G., Munoz, B., Mkocha, H., . . . Lammie, P. J. (2012). CT694 and pgp3 as serological tools for monitoring trachoma programs. *PLoS Negl Trop Dis, 6*(11), e1873. doi:10.1371/journal.pntd.0001873

Gwyn, S., Cooley, G., Goodhew, B., Kohlhoff, S., Banniettis, N., Wiegand, R., & Martin, D. L. (2017). Comparison of platforms for testing antibody responses against the Chlamydia trachomatis antigen Pgp3. *American Journal of Tropical Medicine and Hygiene*. doi:10.4269/ajtmh.17-0292

Gwyn, S., Mitchell, A., Dean, D., Mkocha, H., Handali, S., & Martin, D. L. (2016). Lateral flow-based antibody testing for Chlamydia trachomatis. *Journal of Immunological Methods*. doi:10.1016/j.jim.2016.05.008

Gwyn, S., Mkocha, H., Randall, J. M., Kasubi, M., & Martin, D. L. (2019). Optimization of a rapid test for antibodies to the Chlamydia trachomatis antigen Pgp3. *Diagn Microbiol Infect Dis, 93*(4), 293-298. doi:10.1016/j.diagmicrobio.2018.11.001

Gwyn, S. E., Xiang, L., Kandel, R. P., Dean, D., Gambhir, M., & Martin, D. L. (2018). Prevalence of Chlamydia Trachomatis-Specific antibodies before and after mass drug administration for trachoma in community-wide surveys of four communities in Nepal. *American Journal of Tropical Medicine and Hygiene*. doi:10.4269/ajtmh.17-0102

Martin, D. L., Bid, R., Sandi, F., Goodhew, E. B., Massae, P. A., Lasway, A., . . . Solomon, A. W. (2015). Serology for Trachoma Surveillance after Cessation of Mass Drug Administration. *PLoS Neglected Tropical Diseases*. doi:10.1371/journal.pntd.0003555

MeasuringU. (2018). Confidence Interval Calculator for a Completion Rate. Retrieved from https://measuringu.com/wald/

Migchelsen, S. J., Sepúlveda, N., Martin, D. L., Cooley, G., Gwyn, S., Pickering, H., . . . Roberts, C. H. (2017). Serology reflects a decline in the prevalence of trachoma in two regions of the Gambia. *Scientific Reports*. doi:10.1038/s41598-017-15056-7

Pinsent, A., Solomon, A. W., Bailey, R. L., Bid, R., Cama, A., Dean, D., . . . White, M. T. (2018). The utility of serology for elimination surveillance of trachoma. *Nat Commun, 9*(1), 5444. doi:10.1038/s41467-018-07852-0

Solomon, A. W., Pavluck, A. L., Courtright, P., Aboe, A., Adamu, L., Alemayehu, W., . . . Foster, A. (2015). The Global Trachoma Mapping Project: Methodology of a 34-Country Population-Based Study. *Ophthalmic Epidemiol, 22*(3), 214-225. doi:10.3109/09286586.2015.1037401

West, S. K., Munoz, B., Weaver, J., Mrango, Z., Dize, L., Gaydos, C., . . . Martin, D. L. (2016). Can We Use Antibodies to Chlamydia trachomatis as a Surveillance Tool for National Trachoma Control Programs? Results from a District Survey. *PLoS Neglected Tropical Diseases*. doi:10.1371/journal.pntd.0004352

Wiegand, R. E., Cooley, G., Goodhew, B., Banniettis, N., Kohlhoff, S., Gwyn, S., & Martin, D. L. (2018). Latent class modeling to compare testing platforms for detection of antibodies against the Chlamydia trachomatis antigen Pgp3. *Scientific Reports*. doi:10.1038/s41598-018-22708-9

World Health Organization. (2016a). WHO | Epidemiological situation. Retrieved April 17, 2019, from WHO website: https://www.who.int/trachoma/epidemiology/en/

World Health Organization. (2016b). *Eliminating trachoma: accelerating towards 2020*. *29*(93), 1–28. Retrieved from http://www.trachomacoalition.org/GET2020/

Tables and Figures

Figure 1. Data Analysis Flowchart

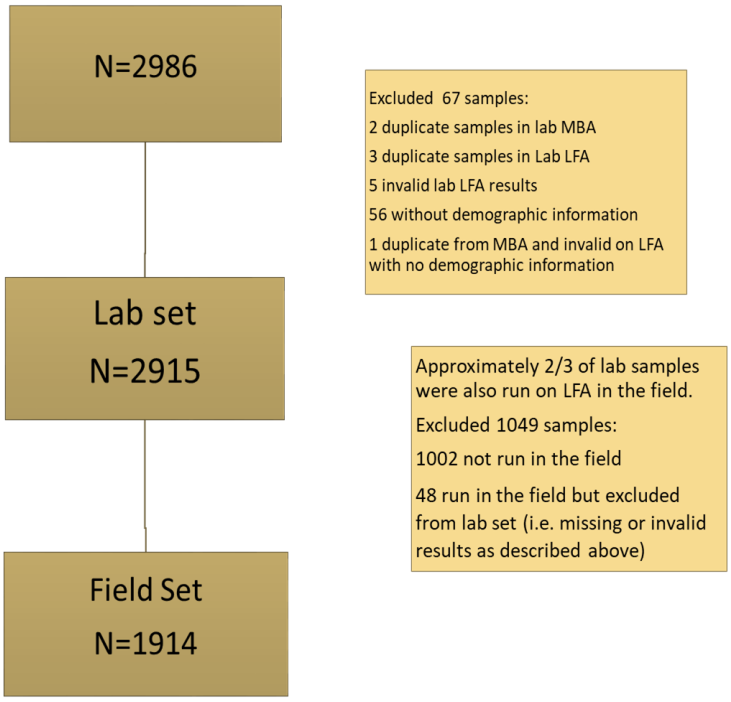


Figure 2. Heatmap of MFI-BG results by Antigen

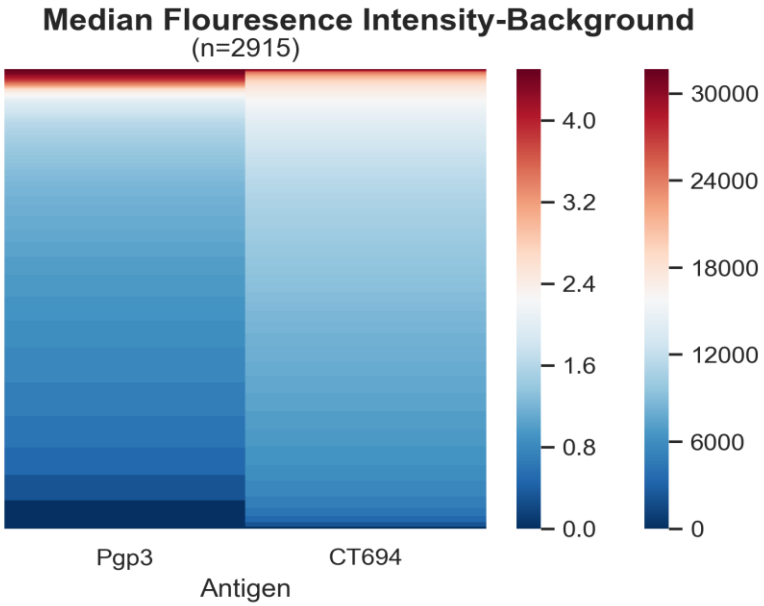
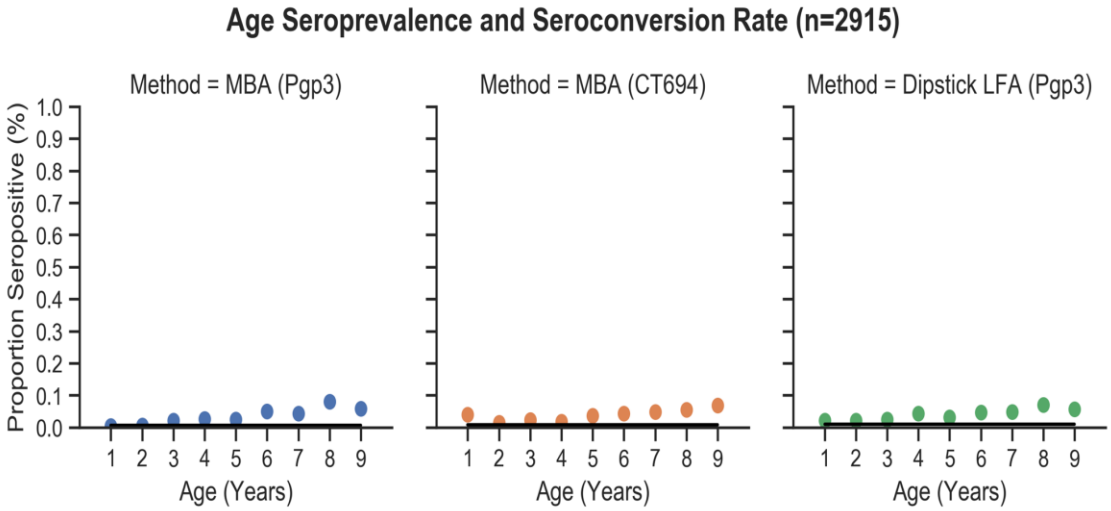


Figure 3. Age Seroprevalence and Seroconversion Rates by Antigen



Appendix

Age Specific Seroprevalence for Lab Dataset (n=2915)

Age	n	Pgp3 (%)	CT694 (%)	Dipstick LFA (%)
1	277	1 (0.4)	1 (0.4)	6 (2.2)
2	279	2 (0.7)	4 (0.1)	6 (2.2)
3	341	7 (2.1)	8 (0.2)	8 (2.3)
4	343	9 (2.6)	6 (0.2)	15 (4.4)
5	376	9 (2.4)	14 (3.7)	12 (3.2)
6	323	16 (5.0)	14 (4.3)	15 (4.6)
7	329	14 (4.3)	16 (4.9)	16 (4.9)
8	308	25 (8.1)	17 (5.5)	22 (7.1)
9	339	20 (5.9)	23 (6.8)	19 (5.6)
All Ages (AVG)		103 (3.5)	103 (3.5)	119 (4.1)

Age Specific Seroprevalence for Field Dataset (n=1914)

Age	n	Pgp3 (%)	CT694 (%)	Dipstick LFA (%)	Cassette LFA (%)
1	179	1 (0.6)	1 (0.6)	5 (2.8)	3 (1.7)
2	190	1 (0.5)	3 (1.6)	6 (3.2)	4 (2.1)
3	223	7 (3.1)	8 (3.6)	7 (3.1)	14 (6.3)
4	230	7 (3.0)	3 (1.3)	12 (5.2)	13 (5.7)
5	255	4 (1.6)	9 (3.5)	7 (2.7)	9 (3.5)
6	203	9 (4.4)	7 (3.4)	10 (4.9)	8 (3.9)
7	255	10 (3.9)	11 (4.3)	9 (3.5)	16 (6.3)
8	187	14 (7.5)	8 (4.3)	12 (6.4)	13 (7.0)
9	232	16 (6.9)	18 (7.8)	13 (5.6)	20 (7.0)
All Ages (AVG)		69 (3.6)	68 (3.6)	81 (4.2)	100 (5.2)

SAS Code for Data Analysis

```
/*Jessica Randall
Graduate Research Assistant
Centers for Disease Control and Prevention*/

/*****************************************************/
libname t "H:\Practicum_Thesis\TOGO\TogoData\Datasets";
ods graphics on;

/*create dummy variables for positivity and negativity by antigen*/
data work.lab;
set t.lab;
label Pgp3pos="Positive by Pgp3, 1=positive 0=negative"
          Ct694pos="Positive by CT694, 1=positive, 0=negative"
          ;
run;
```

```
data work.field;
set t.field;
label Pgp3pos="Positive by Pgp3, 1=positive 0=negative"
         Ct694pos="Positive by CT694, 1=positive, 0=negative"
         ;
run;

/*check contents of created datasets*/
proc contents data=work.lab;
run;
proc contents data=work.field;
run;

************************************;
*demographics*;
proc freq data=work.lab;
tables age sex*eu age*eu sex*eu;
run;
proc freq data=work.field;
tables age sex*eu age*eu sex*eu;
run;
****************************************;
*Lab data*
*seroprevalence*;
PROC LOGISTIC DATA=work.lab PLOTS(ONLY)=ROC simple alpha=0.05;
        class age;
    MODEL Pgp3pos (EVENT='1') = age /ctable ridging=none; *where 1 is
seropositivity by Pgp3;
        OUTPUT OUT=work.labp P=Pgp3phat XBETA=logitp ;
RUN;
/*CT694*/
PROC LOGISTIC DATA=work.lab PLOTS(ONLY)=ROC simple alpha=0.05;
    class age;
        MODEL Ct694pos (EVENT='1') = age/ ctable ridging=none; *where 1
is seropositivity by CT694;
        OUTPUT OUT=work.labc P=CT694phat XBETA=logitc;
RUN;
/*Dipstick LFA*/
PROC LOGISTIC DATA=work.lab PLOTS(ONLY)=ROC simple alpha=0.05;
        class age;
        MODEL LabLFA (EVENT='1') = age/ctable ridging=none; *where 1 is
seropositivity by CT694;
        OUTPUT OUT=work.labl P=LabLFAphat XBETA=logitl ;
RUN;
/*merge work.labp, work.labc, and work.labl and export by CSV*/
/*from CSV, graph these in python to graph*/
proc sort data=work.labp;
by ID;
run;
proc sort data=work.labc;
by ID;
run;
proc sort data=work.labl;
by ID;
run;
```

```
data work.phats (drop=agerand pgp3 ct694 logpgp3 logct694);
merge work.labp work.labc work.labl;
by ID;
run;
proc sort data=work.phats;
by age;
run;

/*get the mean predicted rates of positivity for each age and antigen*/
title 'Age Seroprevalence by each Antigen';
proc means data=work.phats alpha=0.05;
var Pgp3phat CT694phat  lablfaphat;
by age;
run;
title;

/*Field Data*/
*seroprevalence*;
PROC LOGISTIC DATA=work.field PLOTS(ONLY)=ROC simple alpha=0.05;
        class age;
    MODEL Pgp3pos (EVENT='1') = age /ctable ridging=none; *where 1 is
seropositivity by Pgp3;
        OUTPUT OUT=work.fieldp P=Pgp3phat2 XBETA=logit1 ;
RUN;
/*CT694*/
PROC LOGISTIC DATA=work.field PLOTS(ONLY)=ROC simple alpha=0.05;
    class age;
        MODEL Ct694pos (EVENT='1') = age/ ctable ridging=none; *where 1
is seropositivity by CT694;
        OUTPUT OUT=work.fieldc P=CT694phat2 XBETA=logit2;
RUN;
/*Dipstick LFA*/
PROC LOGISTIC DATA=work.field PLOTS(ONLY)=ROC simple alpha=0.05;
        class age;
        MODEL LabLFA (EVENT='1') = age/ctable ridging=none; *where 1 is
seropositivity by CT694;
        OUTPUT OUT=work.fieldl P=LabLFAphat2 XBETA=logit3 ;
RUN;
/*Casette LFA*/
PROC LOGISTIC DATA=work.field PLOTS(ONLY)=ROC simple alpha=0.05;
        class age;
        MODEL FieldLFA (EVENT='1') = age/ctable ridging=none; *where 1
is seropositivity by CT694;
        OUTPUT OUT=work.fieldlf P=fieldLFAphat XBETA=logit4 ;
RUN;
/*merge*/
proc sort data=work.fieldp;
by ID;
run;
proc sort data=work.fieldc;
by ID;
run;
proc sort data=work.fieldl;
by ID;
run;
```

```
proc sort data=work.fieldlf;
by ID;
run;
data work.phatsf (drop=agerand pgp3 ct694 logpgp3 logct694);
merge work.fieldp work.fieldc work.fieldl work.fieldlf;
by ID;
run;
proc sort data=work.phatsf;
by age;
run;

/*get the mean predicted rates of positivity for each age and antigen*/
title 'Age Seroprevalence by each Antigen';
proc means data=work.phatsf alpha=0.05;
var Pgp3phat2 CT694phat2  lablfaphat2 fieldlfaphat;
by age;
run;
title;

*****************************************;
/*agreement*/
proc freq data=work.lab;
tables Pgp3pos*Ct694pos labLFA*Pgp3pos labLFA*Ct694pos/exact agree
nocum;
run;

/*save work datasets as permanent sets and export*/
data t.lab;
set work.lab;
run;

data t.field;
set work.field;
run;
```

Python Code for Data Visualization is attached as a Supplemental file