

**HARNESSING ANTIBODY KINETICS TO IMPROVE EPIDEMIOLOGIC  
INFERENCE: CASE STUDIES IN CHOLERA AND SARS-COV-2**

by  
Forrest Kirby Jones

A dissertation submitted to Johns Hopkins University in conformity with the  
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland  
April 2022

©2022 Forrest Kirby Jones  
All rights reserved

## ABSTRACT

Serological surveillance can complement traditional surveillance systems, providing information about exposure to and protection from pathogens at the population-level. In this dissertation, I investigated how longitudinal serum samples could be leveraged to understand the potential of serological surveillance systems for SARS-CoV-2 and *Vibrio cholerae*.

In mid-2020, we measured anti-receptor binding domain (RBD) IgG, IgA, and IgM antibodies using an enzyme-linked immunosorbent assay (ELISA) in longitudinally collected serum samples (<122 days after symptom onset) from 343 PCR confirmed cases of coronavirus disease 2019 (COVID-19). Using samples from 1,548 pre-pandemic controls, we set a threshold to determine seropositivity with perfect specificity. The median time to seroconversion was approximately 12 days for all three isotypes. Anti-RBD IgA and IgM responses were short-lived while IgG responses decayed slowly.

We also tested 305 serum samples from 48 culture confirmed cholera cases (collected 2-1083 days post-infection) and 3 uninfected household contacts in Bangladesh for serological biomarkers using a multiplex bead assay (MBA) (IgG, IgA, and IgM for 11 antigens), vibriocidal assay (Ogawa and Inaba serotypes), and ELISA (IgG and IgA for 2 antigens). While vibriocidal responses often had high initial fold-changes (52 and 50 fold-change on-average), several MBA-measured antibodies demonstrated robust responses with similar or longer half-lives. Combining all MBA antibody measures allowed for accurate

identification of previous cholera infections including a cross-validated AUC of 92% for infections in the past 200 days.

We also tested 248 serum samples from 51 Haitian volunteers vaccinated with killed whole-cell cholera vaccine for the same MBA markers. Both vaccination and infection stimulated anti-Ogawa OSP and anti-Inaba OSP responses. Classification models trained with anti-CT-B, anti-Ogawa OSP, and anti-Inaba OSP IgG measurements to detect individuals infected <200 days prior misclassified recently vaccinated people as recently infected shortly after receiving a second dose. In simulated cross-sectional surveys, we found that measuring additional markers or knowing vaccination status was sufficient to accurately adjust seroincidence estimates.

I demonstrated how serological data can be used to measure incidence of previous infection with SARS-CoV-2 and *V. cholerae*. Integrated approaches for monitoring seroincidence and population-level immunity for more pathogens should be considered to advance surveillance systems.

## **ADVISORS AND READERS**

**Anna P. Durbin, MD, Committee Chair** Professor of International Health

**Andrew S. Azman, PhD, Advisor** Associate Scientist of Epidemiology

**Justin T. Lessler, PhD, Co-Advisor** Professor of Epidemiology

**Christopher D. Heaney, PhD, Reader** Associate Professor of Environmental Health and Engineering

## **ACKNOWLEDGEMENTS**

There are many individuals and groups that this dissertation would have been impossible without.

Jason Harris: The opportunity to be involved in the generation of data during my PhD has provided me countless valuable lessons. Your encouragement and mentorship have been instrumental to becoming the scientist I am today.

Justin Lessler: There are few people that I have learned as much as I have from you. When I chat with PhD students about the program, I find myself frequently saying, "I remember Justin once told me...." You inspire me to always consider what makes a question important and hold myself to a high standard.

Andrew Azman: It would be impossible to thank you enough for being the incredible mentor you are, but I can say that you make me believe in myself. You have instilled a level of confidence in my own ability where I know I can do things that I once thought were impossible. That is a gift that I will carry far beyond the realms of academic epidemiology, and I am eternally grateful.

Our collaborators at MGH, icddr, and Zanmi Lasante have been incredible colleagues. Richelle Charles, Firdausi Qadri, and Ed Ryan were instrumental to allow us to generate these serological data, and all provided critical feedback on this dissertation. Additionally, we thank the study participants who gave their time, consent, and samples for the advancement of science. The trust that has been placed in me to analyze and report these data means a great deal to me and motivates me to strive to do the most with them.

The ID Dynamics Group and the Surveillance and Outbreak Response Team (SORT) are two groups of amazing people I am so proud to have been able to call myself a part of. I came to Hopkins to work with the ID Dynamics group and have learned something important every meeting. Amy Wesolowski was particularly impactful, having advised me through the maze of graduate school and provided exciting opportunities for research. Emily Gurley also has been heavily involved in my PhD experience, having continually put efforts into making SORT a teaching ground for public health practice.

I would like to thank the Epidemiology Department for all the resources and training I have received, especially during the COVID-19 pandemic. To be a part of this department when the world's eyes have been on us has been the privilege of a lifetime. I truly appreciate the Epidemiology Department's willingness and flexibility to let me take the time to focus on COVID-19 even though it was not part of my original plan for completing my dissertation. Additionally, I must thank the Biostatistics Department and Elizabeth Colantuoni, my MHS advisor, for providing additional perspective during my degree and the training to consider new statistical methodologies for approaching complex problems.

I am grateful to all the other students in my PhD cohort as well as the many others I have befriended. Graduate school has been a journey made exciting by those who undertook it with me. Though the COVID-19 pandemic limited the amount of time we spent together during the second half of my degree, I am grateful for the relationships I have built during this program which I am confident will last much longer than the time we spent in school together.

Lastly, the understanding, kindness, and love of my family and friends has been the foundation on which I built my graduate education. Thank you Wanda and Joe for always taking a keen interest in my education and playing an important role to get me where I am today. Thank you Marcia for our long conversations that I never want to end. Thank you Tessa for often being much wiser than me despite being younger: I learn so much from you. Thank you Mom for giving me both a true love of numbers and a heart of compassion for others. Thank you Dad for pushing me to go further, but remembering to have fun along the way.

## TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>ii</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>iv</b>
<b>LIST OF TABLES</b> .....	<b>x</b>
<b>LIST OF FIGURES</b> .....	<b>xii</b>
<b>CHAPTER 1: Introduction</b> .....	<b>1</b>
<b>References</b> .....	<b>4</b>
<b>CHAPTER 2: Persistence and decay of human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in COVID-19 patients</b> .....	<b>6</b>
<b>Introduction</b> .....	<b>7</b>
<b>Results</b> .....	<b>9</b>
Study cohorts .....	9
Kinetics of anti-SARS-CoV-2 RBD antibody responses.....	10
Accuracy of RBD antibodies for identifying recent SARS-CoV-2 infection.....	10
Combining multiple isotype measurements to improve accuracy .....	11
Estimation of time to seroconversion and seroreversion for each isotype .....	12
Association between RBD responses and the development of neutralizing antibodies targeting the S protein.....	13
Evaluation of cross-reactivity with other coronaviruses.....	13
Comparison of plasma responses to dried blood spots (DBS) .....	14
<b>Discussion</b> .....	<b>14</b>
<b>Materials and Methods</b> .....	<b>19</b>
Study design .....	19
Sample collection.....	20
Dried blood spots (DBS).....	21

Enzyme-linked immunosorbent assay (ELISA) .....	21
Pseudovirus neutralization assay .....	23
Statistical analysis.....	23
Single isotype thresholds.....	23
Random forest classification models .....	24
Analysis of time to seroconversion and seroreversion.....	24
<b>Acknowledgments.....</b>	<b>25</b>
<b>References.....</b>	<b>26</b>
<b>CHAPTER 3: Identifying recent cholera infections using a multiplex bead</b>	
<b>serological assay.....</b>	<b>39</b>
<b>Introduction .....</b>	<b>40</b>
<b>Methods .....</b>	<b>43</b>
Study Population.....	43
Serological testing and data processing .....	45
Statistical Analysis .....	47
<b>Results .....</b>	<b>49</b>
Description of individuals and timing of samples .....	49
Kinetics of biomarkers in confirmed cholera cases .....	50
Identification of recent infections with cross-sectional serologic	
measurements .....	55
Multiplex bead assay panel simplification .....	58
<b>Discussion.....</b>	<b>62</b>
<b>Acknowledgments.....</b>	<b>65</b>
<b>References.....</b>	<b>66</b>
<b>CHAPTER 4: Conducting cholera serosurveillance in partially vaccinated</b>	
<b>populations .....</b>	<b>75</b>
<b>Introduction .....</b>	<b>76</b>
<b>Methods .....</b>	<b>79</b>
Study Population.....	79
Serological testing and data processing .....	80
Statistical Analyses.....	81



<b>Results .....</b>	<b>85</b>
Study population .....	85
Antibody kinetics of vaccinated volunteers and confirmed cholera cases .....	87
Distinguishing recently infected and vaccinated individuals using serological data.....	91
Misclassification of vaccinated individuals as recently infected .....	92
Simulation of serological surveys and comparison of adjustment strategies.....	94
<b>Discussion.....</b>	<b>97</b>
<b>Acknowledgments.....</b>	<b>100</b>
<b>References.....</b>	<b>101</b>
<b>CHAPTER 5: Conclusion .....</b>	<b>108</b>
<b>References.....</b>	<b>114</b>
<b>APPENDICES .....</b>	<b>119</b>
<b>Appendix A: Supplement to Persistence and decay of human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in COVID-19 patients .....</b>	<b>119</b>
<b>Appendix B: Supplement to Identifying recent cholera infections using a multiplex bead serological assay .....</b>	<b>131</b>
<b>Appendix C: Supplement to Conducting cholera serosurveillance in partially vaccinated populations .....</b>	<b>177</b>
<b>CURRICULUM VITAE .....</b>	<b>195</b>

## LIST OF TABLES

<b>Table 2.1. Individual characteristics of PCR-positive SARS-CoV-2 cases and pre-pandemic controls. ....</b>	<b>34</b>
<b>Table 2.2. Predictive accuracy of individual isotypes for classifying controls and cases across time. ....</b>	<b>37</b>
<b>Table 4.1. Individual characteristics of culture confirmed cholera patients and vaccinated individuals ....</b>	<b>86</b>
<b>Table A.1. Full amino acid sequences of the coronavirus receptor-binding domains (RBDs) used in this study. ....</b>	<b>128</b>
<b>Table A.2. Predictive accuracy of multiple isotypes for classifying controls and cases over time since symptom onset. ....</b>	<b>129</b>
<b>Table A.3. Parametric estimates of median time to seroconversion for each isotype by different patient characteristics. ....</b>	<b>130</b>
<b>Table B.1. Individual characteristics of culture confirmed cholera patients and uninfected household contacts ....</b>	<b>147</b>
<b>Table B.2. Biphasic and Exponential decay model comparison.....</b>	<b>154</b>
<b>Table B.3. Estimated duration of half-life and average fold-change from univariate exponential decay models ....</b>	<b>162</b>
<b>Table B.4. Relative parameter values for univariate exponential decay models including covariates for different serological markers ....</b>	<b>164</b>

<b>Table B.5. Comparison of cross-validated AUC between multiple marker random forest models using 45-day, 120-day, 200-day, and 300-day infection windows .....</b>	<b>172</b>
<b>Table B.6. Comparison of cross-validated AUC between multiplex bead assay IgG multiple marker random forest models using 45-day, 120-day, 200-day, and 300-day infection windows .....</b>	<b>173</b>
<b>Table C.1. Estimated duration of half-life and average fold-change from exponential kinetic models for vaccinees and cases.....</b>	<b>187</b>
<b>Table C.2. Number of samples from vaccinated individuals misclassified as recently infected by infection window.....</b>	<b>192</b>
<b>Table C.3. Average and range of seroincidence estimates in simulated serological surveys .....</b>	<b>193</b>
<b>Table C.4. Average and range of coverage estimates in simulated serological surveys .....</b>	<b>194</b>

## LIST OF FIGURES

<b>Figure 2.1. Measurement of IgG, IgM, IgA against SARS-CoV-2 spike protein receptor binding domain among pre-pandemic controls and PCR positive cases.....</b>	<b>35</b>
<b>Figure 2.2. Parametric and nonparametric model estimates of time to seroconversion and seroreversion for each isotype. ....</b>	<b>36</b>
<b>Figure 2.3. SARS-CoV-2 pseudovirus neutralization antibody titers in symptomatic PCR positive cases and correlation with anti-RBD IgG responses.....</b>	<b>38</b>
<b>Figure 3.1. Multiplex bead assay measurements of IgG, IgA, and IgM against <i>V. cholerae</i> O1 antigens among culture confirmed cholera patients. ....</b>	<b>52</b>
<b>Figure 3.2. Estimated duration of half-life and average fold-change from exponential decay models. ....</b>	<b>54</b>
<b>Figure 3.3. Cross-validated area under the receiver operating characteristic curve (cvAUC) and predictor importance rankings for random forest models trained on MBA markers across infection windows.....</b>	<b>56</b>
<b>Figure 3.4. Comparison of cross-validated AUC across random forest models trained on traditional and MBA serological markers for 45-day, 120-day, 200-day, and 300-day infection windows.....</b>	<b>59</b>

<b>Figure 3.5. Specificity and time-varying sensitivity estimates of random forest models trained with leave-one-out cross-validation for 45-day, 120-day, 200-day, and 300-day infection windows using different cut-offs. ....</b>	<b>61</b>
<b>Figure 4.1. Multiplex bead assay measurements of IgG, IgM, and IgA against CT-B, OSP, and TcpA antigens among Haitian vaccinated volunteers.....</b>	<b>88</b>
<b>Figure 4.2. Estimated duration of half-life and average fold-change from exponential kinetic models for vaccinees and cases.....</b>	<b>90</b>
<b>Figure 4.3. Serological measurements among recently infected, recently vaccinated, and neither recently vaccinated nor infected adults and classification model ability to distinguish between groups.....</b>	<b>93</b>
<b>Figure 4.4. Classification probability of samples by random forest models after vaccination and comparison of strategies for adjustment using simulated serological surveys.....</b>	<b>96</b>
<b>Figure A.1. Number of PCR positive cases with a sample taken during each week since symptom onset. ....</b>	<b>119</b>
<b>Figure A.2. Smooth average measurements of IgG, IgM, and IgA against SARS-CoV-2 spike protein receptor binding domain among PCR positive cases across time. ....</b>	<b>120</b>
<b>Figure A.3. Individual trajectories for 16 randomly selected individuals with 4 or more measurements. ....</b>	<b>121</b>

<b>Figure A.4. Measurements of IgG, IgM, and IgA against SARS-CoV-2 spike protein receptor binding domain among pre-pandemic controls and symptomatic PCR positive cases. ....</b>	<b>122</b>
<b>Figure A.5. Receiver operating characteristic curve from random forest models and isotype contributions. ....</b>	<b>123</b>
<b>Figure A.6. Confusion matrices and out-of-bag error estimates for random forest models. ....</b>	<b>124</b>
<b>Figure A.7. Confusion matrices and out-of-bag error estimates for random forest models with downsampled controls. ....</b>	<b>125</b>
<b>Figure A.8. Measurements of IgG, IgA, and IgM against the RBD of other coronaviruses among pre-pandemic controls and PCR positive cases. ..</b>	<b>126</b>
<b>Figure A.9. Correlation between plasma and dried blood spot measurements (DBS).....</b>	<b>127</b>
<b>Figure B.1. Vibriocidal Ogawa measurements among individuals selected from the SMIC and PIC cohorts .....</b>	<b>140</b>
<b>Figure B.2. Timing of sample collection among culture confirmed cases relative to date of infection .....</b>	<b>141</b>
<b>Figure B.3. Postive control dilution curve and four parameter log-logistic curve fits for <i>Vibrio cholerae</i> O1 antigen multiplex bead assay markers..</b>	<b>142</b>
<b>Figure B.4. Postive control dilution curve and four parameter log-logistic curve fits for additional antigen multiplex bead assay markers.....</b>	<b>143</b>

<b>Figure B.5. Relationship between relative antibody units and median fluorescence intensity for additional antigen multiplex bead assay markers</b>	<b>144</b>
<b>Figure B.6. Relationship between relative antibody units and median fluorescence intensity for non-<i>Vibrio cholerae</i> O1 antigen multiplex bead assay markers</b>	<b>145</b>
<b>Figure B.7. Weights used for random forest models at each infection window</b>	<b>146</b>
<b>Figure B.8. Multiplex bead assay relative antibody unit measurements of IgG, IgM, and IgA against <i>V. cholerae</i> O1 antigens among culture confirmed cholera patients</b>	<b>148</b>
<b>Figure B.9. Multiplex bead assay Net MFI measurements of IgG, IgM, and IgA against additional antigens among culture confirmed cholera patients</b>	<b>149</b>
<b>Figure B.10. Multiplex bead assay relative antibody unit measurements of IgG, IgM, and IgA against additional antigens among culture confirmed cholera patients</b>	<b>150</b>
<b>Figure B.11. Multiplex bead assay Net MFI measurements of IgG, IgM, and IgA against <i>V. cholerae</i> O1 antigens among culture confirmed cholera patients, by age group</b>	<b>151</b>

<b>Figure B.12. Multiplex bead assay Net MFI measurements of IgG, IgM, and IgA against <i>V. cholerae</i> O1 antigens among culture confirmed cholera patients, by infecting serotype .....</b>	<b>152</b>
<b>Figure B.13. Correlation matrix of multiplex bead assay biomarkers .....</b>	<b>153</b>
<b>Figure B.14. Individual-level trajectories of Ogawa OSP IgG .....</b>	<b>156</b>
<b>Figure B.15. Individual-level trajectories of Ogawa OSP IgA .....</b>	<b>157</b>
<b>Figure B.16. Individual-level trajectories of Ogawa OSP IgM .....</b>	<b>158</b>
<b>Figure B.17. Individual-level trajectories of CT-B IgG .....</b>	<b>159</b>
<b>Figure B.18. Individual-level trajectories of CT-B IgA .....</b>	<b>160</b>
<b>Figure B.19. Individual-level trajectories of CT-B IgM .....</b>	<b>161</b>
<b>Figure B.20. Cross-validated receiver operator characteristic curves and permutation importance for random forest models using 45-day, 120-day, 200-day, and 300-day infection window .....</b>	<b>170</b>
<b>Figure B.21. Comparison of cross-validated area under the ROC curve between ensemble and random forest models .....</b>	<b>171</b>
<b>Figure B.22. Comparison of cross-validated AUC across random forest models trained on traditional and MBA serological markers excluding anti-CT-B markers for 45-day, 120-day, 200-day, and 300-day infection windows .....</b>	<b>174</b>
<b>Figure B.23. Marker cutoffs to detect recent infection within the past 120 days .....</b>	<b>175</b>



<b>Figure C.1. Timing of sample collection among Haitian vaccinees relative to date of first dose vaccination .....</b>	<b>181</b>
<b>Figure C.2. Comparison of baseline antibody measurements among Haitian vaccinees and Bangladeshi cholera cases.....</b>	<b>182</b>
<b>Figure C.3. Multiplex bead assay net median fluorescence intensity measurements of IgG, IgM, and IgA against other antigens among Haitian vaccinated volunteers.....</b>	<b>183</b>
<b>Figure C.4. Proportion of individuals with a peak value on a given day, by age group and antibody .....</b>	<b>184</b>
<b>Figure C.5. Posterior density of half-life and average fold-change from exponential kinetic models for anti-OSP, anti-CT-B, and anti-TcpA antibodies among vaccinees and cases .....</b>	<b>185</b>
<b>Figure C.6. Posterior density of half-life and average fold-change from exponential kinetic models for additional antibody measurements among vaccinees and cases.....</b>	<b>186</b>

## CHAPTER 1: Introduction

Public health surveillance systems are instrumental for governments to make data-driven decisions to control infectious diseases. They allow monitoring trends in spread and burden, predicting the trajectories of epidemics, and communicating strategically to the public. Traditionally, case-based surveillance (sometimes known as indicator-based surveillance or clinical surveillance) has been the backbone of infectious disease surveillance systems. However, case-based surveillance typically misses mild and/or asymptomatic infections, often relies on health care seeking and access, may have infrequent laboratory confirmation, and may have biases dependent on the design of reporting systems. Serological surveillance may be able to help address some of these limitations.

Upon infection, many pathogens stimulate the immune system to generate antibodies that last long after the pathogen has disappeared from the body. As these antibodies are specific to the infecting pathogen, their presence may be indicative of prior infection or immunity. Serological surveillance (or serosurveillance), where antibody measurements are systematically collected among participants, can provide complementary information about prior rates of infection and population-level immunity.

The purpose of serosurveillance is to make inference on population-level parameters surrounding transmission, burden, and immunity. This contrasts with measuring antibodies for the purposes of diagnosing past exposure or immunity at the individual-level (which may be fraught with scientific and ethical uncertainty

depending on the pathogen). Often, investigators wish to estimate the incidence of infection for an infectious disease by using serological data to estimate the proportion of the population with antibodies specific to a particular pathogen. Further statistical analysis can use these data to estimate fundamental parameters such as the proportion of cases missed by case-based surveillance systems, the infection fatality ratio, or the basic reproductive number. When a correlate of protection (from either infection or disease) is being measured, the proportion of the population that is immune can be estimated. This can be used to alert authorities to populations which are at risk for outbreaks. This type of strategic information could help forecast risk, improve planning of control measures, or optimize allocation of limited resources prior to a crisis.

To implement serosurveillance, investigators often measure antibodies from blood samples though samples of other bodily fluids (such as saliva) can also be tested. The choice of sample type depends on feasibility of implementation, the accuracy of the test, and acceptability to participants. Sometimes samples have already been collected through routine mechanisms (e.g., during blood donations or emergency department visits) while other investigations actively collect samples from the general population or specific populations of interest (e.g., high risk groups). Sampling may be cross-sectional or longitudinal depending on the study or surveillance system goals, epidemiology of the pathogen, and the assay chosen.

In this dissertation, I investigated the potential for serosurveillance of SARS-CoV-2 and *V. cholerae*. Though biologically unrelated, there are several

important qualities shared between these pathogens with regards to serosurveillance. Infection with either SARS-CoV-2 or *V. cholerae* leads to severe symptoms in a minority of individuals (1, 2), but does stimulate a detectable antibody response (3, 4). Additionally, reinfection can occur for both (5, 6). However, most importantly (when this work had begun in mid-2020), few serological studies or surveillance systems had been established for either pathogen (7–9). As with any pathogen, the degree of antibody waning, individual variability, assay misclassification, and cross-reactive immune responses needed to be considered for both pathogens before serosurveillance could be implemented.

To address some of these challenges, I analyzed longitudinal serological data collected after infection (and vaccination for *V. cholerae*) to describe the course of antibody dynamics and assessed the accuracy of assays to detect previous infection. In Chapter 2, I characterized the persistence and decay of the antibody response to SARS-CoV-2 infection among COVID-19 patients. In Chapter 3, I compared the ability of traditional serological assays with a novel assay to identify *V. cholerae* infection during pre-defined time windows before sample collection. In Chapter 4, I developed strategies for conducting serosurveillance to estimate cholera incidence in partially vaccinated populations. Though the analyses for these two pathogens were done separately, this dissertation highlights the similar challenges posed for monitoring both pathogens and that future serosurveillance efforts might be worth integrating across pathogens.

## References

1. E. J. Nelson, J. B. Harris, J. G. Morris Jr, S. B. Calderwood, A. Camilli, Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nat. Rev. Microbiol.* **7**, 693–702 (2009).
2. X. Chen, Z. Chen, A. S. Azman, X. Deng, R. Sun, Z. Zhao, N. Zheng, X. Chen, W. Lu, T. Zhuang, J. Yang, C. Viboud, M. Ajelli, D. T. Leung, H. Yu, Serological evidence of human infection with SARS-CoV-2: a systematic review and meta-analysis. *Lancet Glob Health.* **9**, e598–e609 (2021).
3. J. E. Bryant, A. S. Azman, M. J. Ferrari, B. F. Arnold, M. F. Boni, Y. Boum, K. Hayford, F. J. Luquero, M. J. Mina, I. Rodriguez-Barraquer, J. T. Wu, D. Wade, G. Vernet, D. T. Leung, Serology for SARS-CoV-2: Apprehensions, opportunities, and the path forward. *Sci Immunol.* **5** (2020), doi:10.1126/sciimmunol.abc6347.
4. H. Clapham, J. Hay, I. Routledge, S. Takahashi, M. Choisy, D. Cummings, B. Grenfell, C. J. E. Metcalf, M. Mina, I. R. Barraquer, H. Salje, C. C. Tam, Seroepidemiologic study designs for determining SARS-COV-2 transmission and immunity. *Emerg. Infect. Dis.* **26**, 1978–1986 (2020).
5. W. E. Woodward, Cholera reinfection in man. *J. Infect. Dis.* **123**, 61–66 (1971).
6. J. Wang, C. Kaperak, T. Sato, A. Sakuraba, COVID-19 reinfection: a rapid systematic review of case reports and case series. *J. Investig. Med.* **69**, 1253–1255 (2021).

7. E. Bendavid, B. Mulaney, N. Sood, S. Shah, R. Bromley-Dulfano, C. Lai, Z. Weissberg, R. Saavedra-Walker, J. Tedrow, A. Bogan, T. Kupiec, D. Eichner, R. Gupta, J. P. A. Ioannidis, J. Bhattacharya, COVID-19 antibody seroprevalence in Santa Clara County, California. *Int. J. Epidemiol.* **50**, 410–419 (2021).
8. B. R. Jackson, D. F. Talkington, J. M. Pruckler, M. D. B. Fouché, E. Lafosse, B. Nygren, G. A. Gómez, G. A. Dahourou, W. R. Archer, A. B. Payne, W. C. Hooper, J. W. Tappero, G. Derado, R. Magloire, P. Gerner-Smidt, N. Freeman, J. Boncy, E. D. Mintz, The Cholera Serosurvey Working Group, Seroepidemiologic survey of epidemic cholera in Haiti to assess spectrum of illness and risk factors for severe disease. *Am. J. Trop. Med. Hyg.* **89**, 654–664 (2013).
9. A. S. Azman, S. A. Lauer, T. R. Bhuiyan, F. J. Luquero, D. T. Leung, S. T. Hegde, J. B. Harris, K. K. Paul, F. Khaton, J. Ferdous, J. Lessler, H. Salje, F. Qadri, E. S. Gurley, *Vibrio cholerae* O1 transmission in Bangladesh: insights from a nationally representative serosurvey. *Lancet Microbe.* **1**, e336–e343 (2020).

## **CHAPTER 2: Persistence and decay of human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in COVID-19 patients**

Anita S. Iyer\*, Forrest K. Jones\*, Ariana Nodoushani\*, Meagan Kelly, Margaret Becke, Damien Slater, Rachel Mills, Erica Teng, Mohammad Kamruzzaman, Wilfredo F. Garcia-Beltran, Michael Astudillo, Diane Yang, Tyler E. Miller, Elizabeth Oliver, Stephanie Fischinger, Caroline Atyeo, A. John Iafrate, Stephen B. Calderwood, Stephen A. Lauer, Jingyou Yu, Zhenfeng Li, Jared Feldman, Blake M. Hauser, Timothy M. Caradonna, John A. Branda, Sarah E. Turbett, Regina C. LaRocque, Guillaume Mellon, Dan H. Barouch, Aaron G. Schmidt, Andrew S. Azman, Galit Alter, Edward T Ryan, Jason B. Harris#, Richelle C. Charles#

\*contributed equally to this work

#contributed equally to this work

### **Citation**

Iyer, A. S., Jones, F. K., Nodoushani, A., Kelly, M., Becker, M., Slater, D., ... & Charles, R. C. (2020). Persistence and decay of human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in COVID-19 patients. *Science immunology*, 5(52), eabe0367.

### **Abstract**

We measured plasma and/or serum antibody responses to the receptor-binding domain (RBD) of the spike (S) protein of SARS-CoV-2 in 343 North American

patients infected with SARS-CoV-2 (93% of which required hospitalization) up to 122 days after symptom onset and compared them to responses in 1548 individuals whose blood samples were obtained prior to the pandemic. After setting seropositivity thresholds for perfect specificity (100%), we estimated sensitivities of 95% for IgG, 90% for IgA, and 81% for IgM for detecting infected individuals between 15 and 28 days after symptom onset. While the median time to seroconversion was nearly 12 days across all three isotypes tested, IgA and IgM antibodies against RBD were short-lived with median times to seroreversion of 71 and 49 days after symptom onset. In contrast, anti-RBD IgG responses decayed slowly through 90 days with only 3 seropositive individuals seroreverting within this time period. IgG antibodies to SARS-CoV-2 RBD were strongly correlated with anti-S neutralizing antibody titers, which demonstrated little to no decrease over 75 days since symptom onset. We observed no cross-reactivity of the SARS-CoV-2 RBD-targeted antibodies with other widely circulating coronaviruses (HKU1, 229 E, OC43, NL63). These data suggest that RBD-targeted antibodies are excellent markers of previous and recent infection, that differential isotype measurements can help distinguish between recent and older infections, and that IgG responses persist over the first few months after infection and are highly correlated with neutralizing antibodies.

## **Introduction**

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 (COVID-19), has spread rapidly around the world since being first identified in Wuhan, China, in December 2019



(1). On March 11th, 2020 the World Health Organization (WHO) declared COVID-19 a pandemic, which surpassed 1 million reported global deaths on September 28th, 2020 (2).

Currently, our understanding of antibody responses following infection with SARS-CoV-2 is limited (3–5). Specifically, we lack detailed descriptions and precise estimates concerning the magnitude and duration of responses, cross-reactivity with other coronaviruses and viral respiratory pathogens, and correlates of protective immunity following infection. A detailed characterization of antibody responses is needed to determine whether antibody-based tests can augment viral detection-based assays in the diagnosis of active or recent infection and to inform the design and interpretation of seroepidemiologic studies.

In this study, we characterize the kinetics and antibody isotype profile to the receptor binding domain (RBD) of the spike (S) protein of SARS-CoV-2 in a longitudinal cohort of North American patients infected with SARS-CoV-2, most of whom were hospitalized for COVID-19, and in pre-pandemic controls. We also examined how well these responses correlated with neutralizing antibody activity directed at the S protein. Additionally, we evaluated the cross-reactivity of these responses with other coronavirus RBDs and characterize assay performance using dried blood spots as an alternative to serum or plasma.

## Results

### Study cohorts

Using an in-house enzyme linked immunosorbent assay (ELISA), we measured anti-RBD (Table A.1) antibody responses in two cohorts: 1) symptomatic patients who tested positive for SARS-CoV-2 by PCR (n = 343) and 2) healthy (n = 1,515) and febrile controls (n = 33) collected prior to the SARS-CoV-2 pandemic. Most SARS-CoV-2 positive cases were severe (93% hospitalized, 53% requiring ICU level care, 13% died), male (62%), and older (median age: 59) (Table 2.1, Figure A.1). Most pre-pandemic controls were younger (median age: 37) and female (66%). Plasma and/or serum was collected at multiple time points for most patients (63%; n=216), with 34% (n=118) having  $\geq 4$  samples. Forty-two percent of cases had a sample collected between 0-7 days after onset of symptoms (n=143), 55% had a sample between 8-14 days (n=189), 48% had a sample between 15-28 days (n=165), 35% had a sample between 29-45 days (n=121), 22% had a sample between 46-60 days (n=76), and 10% had a sample > 60 days (n=35). The last sample was collected 122 days post-symptom onset. Twenty-six (8%) cases were immunosuppressed (e.g., on methotrexate, rituximab, etc.), and we did not expect them to mount a robust immune response.

### *Kinetics of anti-SARS-CoV-2 RBD antibody responses*

If followed for more than 14 days since symptom onset, most cases (92%) had at least one IgG measurement higher than seen among any pre-pandemic control (Figure 2.1). From days 5 to 14, there was a sharp rise in RBD-specific antibodies of all isotypes, and IgG measurements continued to rise until day 25 after the onset of symptoms (Figure A.2A). The population average IgA and IgM responses peaked less than a week earlier than IgG and then declined toward concentrations measured in pre-pandemic samples (Figure A.2). IgG antibody responses also began to wane, but at a slower rate. Among 117 cases with  $\geq 4$  measurements, the individual peak IgM measurement often occurred before that of IgG (before: 55%, simultaneous: 38%) and simultaneously with that of IgA (before: 28%, simultaneous: 53%) (Figure A.3). Among hospitalized patients, the population average trajectory differed little between severity levels; the average IgG concentrations among hospitalized cases admitted to the ICU were higher than hospitalized cases not admitted to the ICU (Figure A.2B). Concentrations of all isotypes were lower among immunosuppressed individuals (Figure A.2C).

### *Accuracy of RBD antibodies for identifying recent SARS-CoV-2 infection*

Each antibody isotype was indicative of infection. The area under the receiver operating curve (AUC) for each antibody isotype increased to above 98% during the period of 15-28 days after symptom onset (Table 2.2). The AUC remained high for IgG (99%) and IgA (98%) after 28 days but began to fall for IgM (93%). Using test cutoffs set to ensure no false positives within the pre-pandemic samples (i.e., 100% within sample specificity), we found that the

sensitivity of IgG antibodies rose from 7% ( $\leq 7$  days) to 95% after 14 days of symptoms. The sensitivity of IgA and IgM rose to 90% and 81% 2-4 weeks post-symptom onset but dropped after 4 weeks to 66% and 44%, respectively.

Through ten-fold cross-validation, we found that the mean specificity for each isotype was 99.9% (fold-specific range: 99.4 - 100%).

### *Combining multiple isotype measurements to improve accuracy*

We found the accuracy of serologic identification of recent infections could be slightly improved by adding measurements of IgM and/or IgA to IgG at the earlier phases of infection (Table A.2, Figure A.4). Using random forest models to flexibly combine measurements of different isotypes, we estimated a cvAUC of 92% for IgG & IgM and 91% for IgG & IgA at 8-14 days post-symptom onset. These models provide an estimate of the contribution of each antibody isotype, as well as an approximation of the maximum predictive value of combined measures of anti-RBD IgG, IgA and IgM responses. While all isotypes contributed nearly equally to identifying recent infection antibody profiles in the early phase of illness, IgG responses were the most indicative of infection 8 or more days after the onset of symptoms (Figure A.5). Using the pre-determined thresholds for seropositivity for each antibody isotype, out of the 357 samples collected during early infection ( $< 14$  days post symptom onset), we were able to correctly identify an additional 19 (5%) cases among the IgG negative samples by adding IgM, 21 (6%) by adding IgA, and 33 (9%) by adding both IgM and IgA. When accounting for class imbalance in the random forest procedure, similar results were obtained (Figure A.6, Figure A.7).

### Estimation of time to seroconversion and seroreversion for each isotype

Using the cutoffs defined earlier, we estimated the distribution of the time required to become seropositive (seroconversion) and return to becoming seronegative (seroreversion). Overall, 324 (94%) individuals had more than 1 measurement for every 28 days of follow-up. Of the 159 cases with samples after 20 days post-symptoms, most had evidence of seroconversion for all isotypes (IgG: 96%, IgM: 88%, IgA: 89%). The estimated median time to seroconversion from symptom onset was comparable across antibody isotype: 10.7 days (95% CI: 9.6-11.9) for IgG, 11.7 days (10.4-13.0) for IgA and 11.9 (10.5-13.4 days) for IgM (Figure 2.2). On average, we estimated the median time to seroconversion among hospitalized patients to be over four days earlier as compared to non-hospitalized patients for all isotypes; men and those aged <65 years also seroconverted more quickly on average (Table A.3).

Of seroconverted cases with samples 46 days post-symptoms or after, most eventually had IgM (45/61) and IgA (30/64) seronegative measurements. The median time to seroreversion for IgM was 48.9 days (95% CI: 43.8 – 55.6), with the first 5% seroreverting by 23.7 days (95% CI: 21.6 – 26.0). We estimated a slightly later median seroreversion time for IgA of 70.5 days (95% CI: 58.5 - 87.5), with the first 5% seroreverting by 27.7 days (95% CI: 22.8–32.9, Figure 2.2). Only 3 of 70 cases had evidence of seroreversion for IgG. All 3 patients who seroreverted for IgG required ICU level care; however 2 of the 3 did not have robust IgG responses (peak IgG measurement < 2 µg/mL, 1 of whom was immunosuppressed).

*Association between RBD responses and the development of neutralizing antibodies targeting the S protein*

We measured pseudoneutralizing antibodies targeting the SARS-CoV-2 S protein in 88 samples from 15 individuals collected between 0 and 75 days post-symptoms (Figure 2.3). These individuals were randomly selected from individuals in the full cohort that had multiple samples collected longitudinally with sufficient volume for the pseudoneutralization assay. Over the course of infection, all individuals tested developed detectable neutralizing antibodies (NAb). NAb titers were correlated with the concentration of anti-RBD IgG ( $r = 0.87$ ). Of note, similar to anti-RBD IgG responses, NAb titers plateaued and remained detectable at later time points despite the more rapid decline of IgA and IgM responses.

*Evaluation of cross-reactivity with other coronaviruses*

We evaluated antibody responses to RBDs derived from spike proteins of endemic human coronaviruses (CoVs) (i.e., HKU1, 229E, OC43, and NL63), severe acute respiratory syndrome coronavirus (SARS-CoV-1) and Middle East Respiratory Syndrome coronavirus (MERS-CoV) (Figure A.8). Antibody responses to the endemic CoVs were comparable between pre-pandemic controls and individuals with COVID-19 at all phases of infection, demonstrating a lack of cross-reactivity. Although a few individuals with SARS-CoV-2 infection had increasing levels of antibodies to endemic CoVs over time, which could be explained by cross-reactive anamnestic responses/ immunologic memory, the majority stayed the same. Thus, overall, we did not observe a detectable cross-

reactive response to the RBDs of the endemic human coronaviruses across the population of individuals infected with SARS-CoV-2. In contrast, we did observe significant cross-reactivity to SARS-CoV-1 RBD in individuals with COVID-19, but no significant cross-reactive responses to the MERS-CoV RBD. Of note, there were three pre-pandemic controls (samples collected prior to October 2019) with IgA cross-reactivity to SARS-CoV-1.

#### *Comparison of plasma responses to dried blood spots (DBS)*

Since DBS could be used in large serosurveys where venous blood may be logistically challenging to collect and process, we also evaluated the assay with simulated dried blood spot eluates in a subset of patients (n= 20 at two timepoints; 40 samples) and pre-pandemic controls (n=20). The anti-RBD IgG DBS measurements had a high degree of linear correlation in both cases and control plasma ( $r = 0.99$ , Figure A.9). While the classification of all samples was the same between DBS and plasma samples (100% classification concordance), values between the two sample types diverged more at low titer values.

### **Discussion**

In this study, we found that antibodies against the RBD region of the S protein were accurate indicators of recent severe SARS-CoV-2 infection. The presence of IgG antibodies targeting SARS-CoV-2 RBD was a highly sensitive (95%) marker of infection after 14 days from onset of illness. This is consistent with a growing body of data which demonstrate that measurement of anti-RBD antibodies can accurately classify individuals recently infected with SARS-CoV-2

(6–9). Because this study was conducted in a large cohort of individuals with known SARS-CoV-2 infection (N=343) and controls (N=1548) it provides a robust measure of the accuracy of anti-RBD antibodies.

These findings also add to emerging evidence on the persistence and decay of antibody responses following SARS-CoV-2 infection. IgM and IgA responses to RBD were short-lived and most individuals seroreverted within two and a half months after the onset of illness. This timing and pattern of seroreversion for IgM fits with findings from other pathogens such as SARS-CoV-1 (10), RSV (11), syphilis (12), Zika Virus (13), Chikungunya Virus (14), and West Nile Virus (15). In contrast, IgG antibodies persisted at detectable levels in patients beyond 90 days after symptom onset, and seroreversion was only observed in a small percentage of individuals. The concentration of these anti-RBD IgG antibodies was also highly correlated with pseudovirus NAb titers, which also demonstrated minimal decay. The observation that IgG and neutralizing antibody responses persist is encouraging and suggests the development of robust systemic immune memory in individuals with severe infection. This is similar to a study that reported on anti-RBD antibodies in 121 North American convalescent plasma donors up to 82 days from symptom onset (16) and a study of 1,197 Icelanders who remained seropositive by 2 pan-IgG SARS-CoV-2 antibody assays 120 days after qPCR diagnosis of SARS-CoV-2 (9). However, these findings differ with other recent studies suggesting a more rapid waning in anti-RBD titers following mild or asymptomatic SARS-CoV-2 infection (17, 18).



RT-PCR based detection of SARS-CoV-2 is sensitive early in the first week after the onset of symptoms (19), and our results suggest that the detection of antibodies against the SARS-CoV-2 RBD by ELISA, even when utilizing all isotypes, is not likely to contribute significantly to the early diagnosis of COVID-19. However, beyond two weeks after symptom onset, supplementing viral detection assays with antibody-based testing methods clearly increases sensitivity in diagnosing recent infection (20, 21), particularly as the sensitivity of RT-PCR for SARS-CoV-2 infection wanes (18). In particular, our results demonstrate that the earlier seroreversion of IgA and IgM responses will be helpful in distinguishing older infections from recent ones. Thus, the measurement of multiple isotypes, taking into account the early decay of IgA and IgM, is likely to be critical in interpreting the results of serosurveys and epidemiologic studies to estimate the time from infection. All considered, these findings suggest clearly defined applications for serologic testing of RBD responses in both clinical and public health/surveillance settings.

Testing for anti-SARS-CoV-2 RBD antibodies can also be applied in seroepidemiologic studies, even in areas of low prevalence, given their excellent specificity and defined kinetics. Variation in the performance of commercial serologic tests and confusion about the role of antibodies as biomarkers of past infection versus protective immunity has led to widespread misperception that antibody testing may be inaccurate (22, 23). In contrast, our study, based on a very large sample of cases and controls, should provide significant confidence in the contribution of serologic measures in public health efforts to improve

epidemiological investigations (24) and to provide high-resolution estimates of infection incidence across geographies and populations. In addition, the lack of cross-reactivity of antibodies to SARS-CoV-2 RBD with common cold coronaviruses provides additional data supporting the specificity of the assay.

One limitation of our study was that our cohort of individuals with SARS-CoV-2 infection was skewed toward adults with severe disease or with risk factors for disease progression. It is important to study the kinetics and in particular the decay of antibody responses in individuals with severe infection for several reasons. First, the magnitude and duration of the responses in individuals with severe infection likely provide an estimate of the upper bounds of the achievable immune response and the development of B cell memory following natural infection. Second, these findings are expected to have significant implications for protective immunity in a population which clearly is vulnerable to poor outcomes when exposed. However, caution is required in generalizing these results to those with less severe infection. Individuals with mild or asymptomatic infection have been shown to develop less robust antibody responses (18), which may lead to false negatives if our proposed assay thresholds are used. Individuals with mild or asymptomatic infection may also serorevert more quickly than symptomatic individuals. The gradation of responses by disease severity has been found in other infections, including SARS-CoV-2 and MERS-CoV infection (25). An association between disease severity and the kinetics of the antibody response is also suggested by our finding that individuals with more severe disease, who required ICU-level care,

seroconverted earlier than individuals who did not require ICU-level support. However, some patients, such as the elderly individuals or diabetics, may have more severe disease because of a poor immune response with potentially lower concentrations of antibodies present.

While anti-RBD antibodies accurately identify individuals with recent SARS-CoV-2 infection, it remains unknown whether these responses are associated with protection against subsequent infection. In many human challenge studies of common cold coronavirus infection, the presence of neutralizing antibodies has been associated with protection against symptomatic infection and decreased viral shedding (5). In addition, in vaccinated rhesus macaques challenged with SARS-CoV-2 infection, neutralizing antibodies directed at the S protein were also a strong correlate of protective immunity (26). Thus, neutralization titers, in the absence of other known markers, have become a de facto immunologic marker of protection against infection pending further investigation. In this context, it is notable that anti-RBD IgG antibodies were strongly correlated with the same neutralizing antibodies that were associated with protection in vaccinated macaques (26). This correlation with neutralizing titers was stronger than observed for other previously tested commercial serologic assays (27), and both anti-RBD and neutralizing antibodies persisted over a 2.5 month follow-up period.

Our results, therefore, provide strong support for the application of anti-RBD antibodies as a marker of recent SARS-CoV-2 infection as well as new and detailed information related to the specificity and decay kinetics of the anti-RBD

responses. The testing approach used meets the CDC's guidelines for serologic testing (28) and has the potential to facilitate accurate diagnosis in clinical settings and the implementation of population-based studies of previous infection globally. While the association between anti-RBD-IgG and neutralizing titers and the persistence of these antibodies at late time points is encouraging, further work is needed to define the optimal antibody-mediated correlates of protective immunity.

## **Materials and Methods**

### *Study design*

We evaluated the magnitude and kinetics of the early human antibody response to the receptor binding domain of SARS-CoV-2 spike protein, with the additional objective of evaluating the specificity and sensitivity of these antibody responses for identifying individuals with recent infection. Thus, we measured antibody concentration in blood samples obtained from confirmed patients with symptomatic SARS-CoV-2 infection and from control individuals whose samples were collected prior to the pandemic. IgG, IgA, and IgM antibody concentrations were measured by ELISA using recombinant SARS-CoV2 RBD in all samples. We focused on the RBD given its importance for viral entry into cells and that antibodies binding to the RBD may play an important role in neutralization. In a subset of samples, neutralizing antibody responses directed against the spike protein were also measured using a lentivirus pseudoneutralization model. From these data, we modeled the classification accuracy for each individual isotype

and combinations of isotypes at different time points, and the temporal dynamics of seroconversion and seroreversion following the onset of symptoms.

### Sample collection

We obtained plasma and/or serum samples, collected for routine clinical care, from individuals with PCR-confirmed SARS-CoV-2 infection presenting, with fever and/or viral respiratory symptoms from March to April 2020 and who met criteria for RT-PCR testing. Testing criteria for SARS-CoV-2 changed over time, but primarily included patients with severe symptoms requiring hospital admission, although those who had other risk factors for disease progression (e.g., were age 60 or older, had diabetes, or were immunocompromised), or who worked or lived in a setting where infection control requirements dictated a need for testing. Additional serum/plasma samples collected September 2015 to December 2019 prior to the SARS-CoV-2 pandemic included healthy adults seen at the MGH Immunization and Travel Clinic prior to travel, patients undergoing routine serology, and patients presenting with other known febrile illnesses. Plasma samples, except for the routine serology samples, were heat-inactivated at 56°C for one hour prior to analysis. Patient demographic information, lab results, and clinical outcomes were extracted from the electronic medical record. Patients were considered immunosuppressed if they had underlying immunosuppressive condition (e.g., HIV with CD4 count less than 200) or were on an immunosuppressive/immunomodulating agent at the time of their admission (e.g., methotrexate, rituximab) All research was approved by the Institutional Review Board for Human Subjects Research at MGH.

### Dried blood spots (DBS)

Seventy-two microliters of single donor, seronegative whole blood collected from sodium heparin tubes (Becton, Dickinson, NJ), was spiked with 8  $\mu$ L heat-inactivated plasma (10% of the whole blood volume) to maintain the relative whole blood composition. Assuming plasma is 50% of the whole blood volume, the spiked plasma was 18.18% of the final plasma volume. Whole blood (40  $\mu$ L) was spotted onto Whatman 903 Protein Saver cards (GE Healthcare, Cardiff, UK) in replicate and allowed to dry overnight at room temperature. Two 6-mm<sup>2</sup> punches from the DBS card (5  $\mu$ L plasma per punch) were placed in 133  $\mu$ L PBS-0.05% Tween 20 pH 7.4 (Sigma-Aldrich, St. Louis, MO) and incubated overnight at 4°C with gentle agitation eluates were then recovered after centrifugation. The total dilution of the spiked plasma in DBS eluate was assumed to be 1:73.15, which accounts for the initial dilution from spiking (1:5.5) and the further dilution during DBS elution (1:13.3)

### Enzyme-linked immunosorbent assay (ELISA)

The ELISA assays measured IgG, IgA, and IgM responses to the receptor binding domain of the spike protein (RBD) from SARS-CoV-2 [GenBank: MN975262], Middle East Respiratory Syndrome (MERS) virus [GenBank: AFY13307.1], SARS-CoV-1 [GenBank: AAP13441.1], and common cold coronaviruses HKU1 [GenBank: AAT98580.1], OC229E [GenBank: AAK32191], OC43 [GenBank:AAT84362], and NL63 [GenBank: AKT07952]. RBD sequences were cloned into pVRC vector followed by expression in mammalian cells Expi293 cells (ThermoFisher Scientific, Waltham,MA) with a C-terminal

streptavidin-binding peptide (SBP)-His8X tag, and purified over TALON resin (Takara, Mountain View, CA) followed by size exclusion chromatography and cleavage of the His tag. RBD-specific antibody concentrations ( $\mu\text{g}/\text{mL}$ ) were quantified using isotype-specific anti-RBD monoclonal antibodies. The RBDs were expressed in Expi293F suspension cells with a C-terminal SBP-His8X tag, and purified using affinity chromatography and then size exclusion chromatography prior to removal of the His tag as described previously (29). Briefly, 384 well Nunc MaxiSorp plates (Invitrogen, Carlsbad, CA) were coated by adding 50  $\mu\text{L}$  of RBD in carbonate buffer (1  $\mu\text{g}/\text{mL}$ ) and incubating for 1 hour at room temperature (RT). Plates were then blocked for 30 min at RT with 5% nonfat milk in tris-buffered saline (TBS). Diluted samples (1:100 in TBS with 5% milk, 0.5% Tween) were added to the plate (25  $\mu\text{L}/\text{well}$ ) and incubated for 1 hour at 37°C with shaking. Serial 4-fold dilutions to 1:6400 were also included for individuals with high titers. At the end of incubation, samples were washed 5 times with 1X high salt TBS. Subsequently, goat anti-human IgA, IgG, and IgM-horseradish peroxidase conjugated secondary antibodies diluted (Jackson ImmunoResearch) at 1:10000 (IgG, IgM) or 1:5000 (IgA) in 5% milk in TBST were added to plates (25  $\mu\text{L}/\text{well}$ ) and incubated at RT with shaking for 30 min followed by 5X 1X high salt TBS washes and a last wash with 1X TBS. Bound secondaries were detected using 1-step Ultra TMB (tetramethylbenzidine; ThermoScientific, Waltham, MA, 25  $\mu\text{L}/\text{well}$ ). Plates were incubated at RT for 5 min in the dark before addition of 2 N sulfuric acid stop solution (25  $\mu\text{L}/\text{well}$ ). The optical density (OD) was read at 450 nm and 570 nm on a plate reader. OD

values were adjusted by subtracting the 570 nm OD from the 450 nm OD. We used a standard curve of the anti-SARS-CoV-2 monoclonal, CR3022 (30), to calculate the concentration of anti-RBD IgG, IgA, and IgM expressed in  $\mu\text{g/mL}$ . Note: For the DBS and plasma comparisons the starting concentration was 1:200.

### *Pseudovirus neutralization assay*

To determine the SARS-CoV-2 neutralization activity of our plasma samples, we used a lentivirus pseudoneutralization model as previously described (26), which is a strong correlate of protective immunity in challenged rhesus macaques (31). We expressed results from this assay as the antibody titer required to neutralize 50% of the SARS-CoV-2 pseudovirus (NT50).

### *Statistical analysis*

#### *Single isotype thresholds*

We first explored how cutoffs of individual isotypes (IgM, IgG and IgA) performed in identifying previously infected individuals. We compared measurements from pre-pandemic controls, with those taken at any time,  $\leq 7$  days, 8-14 days, 15-28 days, and  $>28$  days after the onset of symptoms. We estimated the AUC for each isotype and time period combination and calculated bootstrap 95% confidence intervals. Using the isotype cutoffs defined by the maximum concentration ( $\mu\text{g/mL}$ ) found among the full set of pre-pandemic controls (IgG: 0.57, IgM: 2.63, IgA: 2.02), we estimated sensitivity and bootstrap 95% confidence intervals. We also evaluated how setting a cutoff defined by maximum concentration would affect specificity through ten-fold cross-validation.



### Random forest classification models

We explored how combining multiple isotype-specific responses with random forest classification models, which allows for complex nonlinear interactions between isotypes, performed identifying previously infected individuals. Using a previously described cross-validation procedure (32), we allocated both cases and controls into ten equally sized groups (i.e., folds) and calculated a pooled cross-validated AUC (cvAUC). We also assessed variable importance within these different models using a permutation test-based metric, mean decrease in accuracy. To investigate the impact of class imbalance (i.e., the fact that we had many more negative controls than positives) on our model performance metrics, we investigated the effect of downsampling controls to have the same number as cases on model performance.

### Analysis of time to seroconversion and seroreversion

We limited our analysis to individuals who had at least one measurement for every 28 days of follow-up (i.e., between symptom onset and their last measurement). Measurements above the isotype-specific cut-offs (i.e. the maximum measurements observed among pre-pandemic controls) were considered seropositive, while those below or equal were considered seronegative. For individuals with fluctuations around the pre-defined cutoff (N=6), the time to the first event was used in the analysis. Using individual level interval-censored data, we fitted nonparametric (i.e Turnbull's Estimator) and parametric accelerated failure time models using the icenReg R package (33). All time-to-event data were assumed to be log-normal distributed. Bootstrapped

95% confidence intervals were estimated by sampling individuals with replacement.

All analyses were completed using R (Version 3.6.1) within Rstudio (Version 1.2.5019).

## **Acknowledgments**

*Funding.* This work was supported in part by Centers for Disease Control and Prevention (U01CK000490 to E.T.R, R.C.L., R.C.C, J.B.H, G.M., E.O., S.E.T.), the National Institutes of Health (R01 AI146779 to A.G.S; R01 T32 GM007753 to B.M.H and T.M.C.; R01 AI135115 to A.S.A. and F.K.J. and T32 AI007245 to J.F) and MassCPR grant to A.G.S. *Competing interests.* G.A. has a financial interest in SeromYx, a company developing a platform technology that describes the antibody immune response. J.A.B. has been a consultant for T2 Biosystems, DiaSorin, and Roche Diagnostics. All other authors declare that they have no competing interests. *Data and materials availability.* All data and code needed to reproduce the analysis and assess the conclusions in the paper are present in a GitHub repository (<https://github.com/fjones2222/covid19-serodynamics>) . The data needed to evaluate the conclusions in the paper is present in the paper or the Supplementary Materials. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided

the original work is properly cited. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

## References

1. M. Lipsitch, D. L. Swerdlow, L. Finelli, Defining the Epidemiology of Covid-19 - Studies Needed. *N. Engl. J. Med.* 382, 1194–1196 (2020).
2. WHO Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int/>. Accessed 28th September,2020.
3. T. Zohar, G. Alter, Dissecting antibody-mediated protection against SARS-CoV-2. *Nat. Rev. Immunol.* 20, 392–394 (2020).
4. M. Döhla, C. Boesecke, B. Schulte, C. Diegmann, E. Sib, E. Richter, M. Eschbach-Bludau, S. Aldabbagh, B. Marx, A.-M. Eis-Hübinger, R. M. Schmithausen, H. Streeck, Rapid point-of-care testing for SARS-CoV-2 in a community screening setting shows low sensitivity. *Public Health* 182, 170–172 (2020).
5. T. Huang, B. Garcia-Carreras, M. D. T. Hitchings, B. Yang, L. C. Katzelnick, S. M. Rattigan, B. A. Borgert, C. A. Moreno, B. D. Solomon, L. Trimmer-Smith, V. Etienne, I. Rodriguez-Barraquer, J. Lessler, H. Salje, D. S. Burke, A. Wesolowski, D. A. T. Cummings, A systematic review of antibody mediated immunity to coronaviruses: Kinetics, correlates of protection, and association with severity. *Nat. Commun.* 11, 4704 (2020).

6. L. Premkumar, B. Segovia-Chumbez, R. Jadi, D. R. Martinez, R. Raut, A. Markmann, C. Cornaby, L. Bartelt, S. Weiss, Y. Park, C. E. Edwards, E. Weimer, E. M. Scherer, N. Rouphael, S. Edupuganti, D. Weiskopf, L. V. Tse, Y. J. Hou, D. Margolis, A. Sette, M. H. Collins, J. Schmitz, R. S. Baric, A. M. de Silva, The receptor binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Sci. Immunol.* 5, eabc8413 (2020).
7. L. Ren, G. Fan, W. Wu, L. Guo, Y. Wang, X. Li, C. Wang, X. Gu, C. Li, Y. Wang, G. Wang, F. Zhou, Z. Liu, Q. Ge, Y. Zhang, H. Li, L. Zhang, J. Xu, C. Wang, J. Wang, B. Cao, Antibody Responses and Clinical Outcomes in Adults Hospitalized with Severe COVID-19: A Post hoc Analysis of LOTUS China Trial. *Clin. Infect. Dis.* ciaa1247 (2020).
8. K. M. McAndrews, D. P. Dowlatshahi, J. Dai, L. M. Becker, J. Hensel, L. M. Snowden, J. M. Leveille, M. R. Brunner, K. W. Holden, N. S. Hopkins, A. M. Harris, J. Kumpati, M. A. Whitt, J. J. Lee, L. L. Ostrosky-Zeichner, R. Papanna, V. S. LeBleu, J. P. Allison, R. Kalluri, Heterogeneous antibodies against SARS-CoV-2 spike receptor binding domain and nucleocapsid with implications for COVID-19 immunity. *JCI Insight* 5, 142386 (2020).
9. D. F. Gudbjartsson, G. L. Norddahl, P. Melsted, K. Gunnarsdottir, H. Holm, E. Eythorsson, A. O. Arnthorsson, D. Helgason, K. Bjarnadottir, R. F. Ingvarsson, B. Thorsteinsdottir, S. Kristjansdottir, K. Birgisdottir, A. M. Kristinsdottir, M. I. Sigurdsson, G. A. Arnadottir, E. V. Ivarsdottir, M. Andresdottir, F. Jonsson, A. B. Agustsdottir, J. Berglund, B. Eiriksdottir, R.

- Fridriksdottir, E. E. Gardarsdottir, M. Gottfredsson, O. S. Gretarsdottir, S. Gudmundsdottir, K. R. Gudmundsson, T. R. Gunnarsdottir, A. Gylfason, A. Helgason, B. O. Jensson, A. Jonasdottir, H. Jonsson, T. Kristjansson, K. G. Kristinsson, D. N. Magnusdottir, O. T. Magnusson, L. B. Olafsdottir, S. Rognvaldsson, L. le Roux, G. Sigmundsdottir, A. Sigurdsson, G. Sveinbjornsson, K. E. Sveinsdottir, M. Sveinsdottir, E. A. Thorarensen, B. Thorbjornsson, M. Thordardottir, J. Saemundsdottir, S. H. Kristjansson, K. S. Josefsdottir, G. Masson, G. Georgsson, M. Kristjansson, A. Moller, R. Palsson, T. Gudnason, U. Thorsteinsdottir, I. Jonsdottir, P. Sulem, K. Stefansson, Humoral Immune Response to SARS-CoV-2 in Iceland. *N. Engl. J. Med.* •••, (2020).
10. Z. He, Q. Dong, H. Zhuang, S. Song, G. Peng, G. Luo, D. E. Dwyer, Kinetics of severe acute respiratory syndrome (SARS) coronavirus-specific antibodies in 271 laboratory-confirmed cases of SARS. *Clin. Diagn. Lab. Immunol.* **11**, 792–794 (2004).
11. R. C. Welliver, T. N. Kaul, T. I. Putnam, M. Sun, K. Riddlesberger, P. L. Ogra, The antibody response to primary and secondary infection with respiratory syncytial virus: kinetics of class-specific responses. *J. Pediatr.* **96**, 808–813 (1980).
12. V. Paparizos, A. Tsimpidakis, E. Nicolaidou, E. Daskalakis, E. Paparizou, V. Vasalou, G. Vrioni, D. Rigopoulos, Duration of anti-treponemal immunoglobulin M seroreversion after successful syphilis treatment in HIV-

- positive and -negative patients. *International Journal of STD & AIDS*. **32** (2021), pp. 523–527.
13. M. Stone, S. Bakkour, M. C. Lanteri, D. Brambilla, G. Simmons, R. Bruhn, Z. Kaidarova, T.-H. Lee, J. Orlando Alsina, P. C. Williamson, S. A. Galel, L. L. Pate, J. M. Linnen, S. Kleinman, M. P. Busch, NHLBI Recipient Epidemiology Donor Evaluation Study REDS-III Program, Zika virus RNA and IgM persistence in blood compartments and body fluids: a prospective observational study. *Lancet Infect. Dis.* **20**, 1446–1456 (2020).
14. D. Ramful, S. Sampéris, X. Fritel, A. Michault, M.-C. Jaffar-Bandjee, O. Rollot, B. Boumahni, P. Gérardin, Antibody kinetics in infants exposed to Chikungunya virus infection during pregnancy reveals absence of congenital infection. *J. Infect. Dis.* **209**, 1726–1730 (2014).
15. H. E. Prince, W. R. Hogrefe, Detection of West Nile virus (WNV)-specific immunoglobulin M in a reference laboratory setting during the 2002 WNV season in the United States. *Clin. Diagn. Lab. Immunol.* **10**, 764–768 (2003).
16. F. A. Ania Wajnberg, Adolfo Firpo, Deena Altman, Mark Bailey, Mayce Mansour, Meagan McMahon, Philip Meade, Damodara Rao Mendu, Kimberly Muellers, Daniel Stadlbauer, Kimberly Stone, Shirin Strohmeier, Judith Aberg, David Reich, Florian Krammer, Carlos Cordon-Cardo, SARS-CoV-2 infection induces robust, neutralizing antibody responses that are stable for at least three months. medRxiv (2020), <https://www.medrxiv.org/content/10.1101/2020.07.14.20151126v1>

17. F. Javier Ibarondo, A. Jennifer, Fulcher, David Goodman-Meza, Julie Elliott, Christian Hofmann, Mary A. Hausner, Kathie G. Ferbas, Nicole H. Tobin, Grace M. Aldrovandi, Otto O. Yang, Rapid Decay of Anti-SARS-CoV-2 Antibodies in Persons with Mild Covid-19. *N. Engl. J. Med.* (2020).
18. Q.-X. Long, X.-J. Tang, Q.-L. Shi, Q. Li, H.-J. Deng, J. Yuan, J.-L. Hu, W. Xu, Y. Zhang, F. J. Lv, K. Su, F. Zhang, J. Gong, B. Wu, X.-M. Liu, J.-J. Li, J.-F. Qiu, J. Chen, A.-L. Huang, Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* 26, 1200–1204 (2020).
19. L. M. Kucirka, S. A. Lauer, O. Laeyendecker, D. Boon, J. Lessler, Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction-Based SARS-CoV-2 Tests by Time Since Exposure. *Ann. Intern. Med.* 173, 262–267 (2020).
20. L. Guo, L. Ren, S. Yang, M. Xiao, D. Chang, F. Yang, C. S. Dela Cruz, Y. Wang, C. Wu, Y. Xiao, L. Zhang, L. Han, S. Dang, Y. Xu, Q.-W. Yang, S.-Y. Xu, H.-D. Zhu, Y.-C. Xu, Q. Jin, L. Sharma, L. Wang, J. Wang, Profiling Early Humoral Response to Diagnose Novel Coronavirus Disease (COVID-19). *Clin. Infect. Dis.* 71, 778–785 (2020).
21. S. E. F. Yong, D. E. Anderson, W. E. Wei, J. Pang, W. N. Chia, C. W. Tan, Y. L. Teoh, P. Rajendram, M. P. H. S. Toh, C. Poh, V. T. J. Koh, J. Lum, N. M. Suhaimi, P. Y. Chia, M. I. Chen, S. Vasoo, B. Ong, Y. S. Leo, L. Wang, V. J. M. Lee, Connecting clusters of COVID-19: An epidemiological and serological investigation. *Lancet Infect. Dis.* 20, 809–815 (2020).

22. Antibody Test, Seen as Key to Reopening Country, Does Not Yet Deliver in The New York Times, (Published April 19, 2020).
23. Antibody tests for Covid-19 wrong up to half the time, CDC in CNN. (May 26, 2020).
24. Q. B. Zhen Zhang M, Shisong Fang, Lan Wei, Xin Wang, Jianfan He, Yongsheng Wu BS1, Xiaojian Liu MMed, Wei Gao MMed, Renli Zhang, Wenfeng Gong, Qiru Su, Andrew S Azman, Justin Lessler, Xuan Zou. (2020). Insights into the practical effectiveness of RT-PCR testing for SARS-CoV-2 from serologic data, a cohort study (2020). Published online September 08, 2020. <https://doi.org/>
25. Ko. Jae-Hoon, A. Marcel, Müller, Hyeri Seok, Ga Eun Park, Ji Yeon Lee, Sun Young Cho, Young Eun H, Jin Yang Baek, So Hyun Kim, Ji-Man Kang, Yae-Jean Kim, Ik Joon Jo, Chi Ryang Chung, Myong-Joon Hahn, Christian Drosten, Cheol-In Kang, a Doo Ryeon Chung, Jae-Hoon Song, Eun-Suk Kang, and Kyong Ran Peck, Serologic responses of 42 MERS-coronavirus-infected patients according to the disease severity. *Diagn. Microbiol. Infect. Dis.* 89, 106–111 (2017).
26. Chandrashekar, J. Liu, A. J. Martinot, K. McMahan, N. B. Mercado, L. Peter, L. H. Tostanoski, J. Yu, Z. Maliga, M. Nekorchuk, K. Busman-Sahay, M. Terry, L. M. Wrijil, S. Ducat, D. R. Martinez, C. Atyeo, S. Fischinger, J. S. Burke, M. D. Slein, L. Pessaint, A. Van Ry, J. Greenhouse, T. Taylor, K. Blade, A. Cook, B. Finneyfrock, R. Brown, E. Teow, J. Velasco, R. Zahn, F. Wegmann, P. Abbink, E. A. Bondzie, G. Dagotto, M. S. Gebre, X. He, C.



- Jacob-Dolan, N. Kordana, Z. Li, M. A. Lifton, S. H. Mahrokhian, L. F. Maxfield, R. Nityanandam, J. P. Nkolola, A. G. Schmidt, A. D. Miller, R. S. Baric, G. Alter, P. K. Sorger, J. D. Estes, H. Andersen, M. G. Lewis, D. H. Barouch, SARS-CoV-2 infection protects against rechallenge in rhesus macaques. *Science* 369, 812–817 (2020).
27. R. A. D. Elena Criscuolo, Marta Strollo, Serena Rolla, Alessandro Ambrosi, Massimo Locatelli, Roberto Burioni, Nicasio Mancini, Massimo Clementi, Nicola Clementi, Poor correlation between antibody titers and neutralizing activity in sera from SARS-CoV-2 infected subjects. *medRxiv* (2020), <https://www.medrxiv.org/content/10.1101/2020.07.10.20150375v1>
28. Centers for Disease Control and Prevention, Interim Guidelines for COVID-19 Antibody Testing at <https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/antibody-tests-guidelines.html>. (2020). (2020). Accessed 27th, July 2020. (2020).
29. M. Norman, T. Gilboa, A. F. Ogata, A. M. Maley, L. Cohen, Y. Cai, J. Zhang, J. E. Feldman, B. M. Hauser, T. M. Caradonna, B. Chen, A. G. Schmidt, G. Alter, R. C. Charles, E. T. Ryan, D. R. Walt, Ultra-Sensitive High-Resolution Profiling of Anti-SARS-CoV-2 Antibodies for Detecting Early Seroconversion in COVID-19 Patients. *medRxiv*, (2020), <https://www.medrxiv.org/content/10.1101/2020.04.28.20083691v1>, Published online, May 02, 2020.
30. C. Wang, W. Li, D. Drabek, N. M. A. Okba, R. van Haperen, A. D. M. E. Osterhaus, F. J. M. van Kuppeveld, B. L. Haagmans, F. Grosveld, B.-J.

- Bosch, Publisher Correction: A human monoclonal antibody blocking SARS-CoV-2 infection. *Nat. Commun.* 11, 2511 (2020).
31. J. Yu, L. H. Tostanoski, L. Peter, N. B. Mercado, K. McMahan, S. H. Mahrokhian, J. P. Nkolola, J. Liu, Z. Li, A. Chandrashekar, D. R. Martinez, C. Loos, C. Atyeo, S. Fischinger, J. S. Burke, M. D. Slein, Y. Chen, A. Zuiani, F. J. N. Lelis, M. Travers, S. Habibi, L. Pessaint, A. Van Ry, K. Blade, R. Brown, A. Cook, B. Finneyfrock, A. Dodson, E. Teow, J. Velasco, R. Zahn, F. Wegmann, E. A. Bondzie, G. Dagotto, M. S. Gebre, X. He, C. Jacob-Dolan, M. Kirilova, N. Kordana, Z. Lin, L. F. Maxfield, F. Nampanya, R. Nityanandam, J. D. Ventura, H. Wan, Y. Cai, B. Chen, A. G. Schmidt, D. R. Wesemann, R. S. Baric, G. Alter, H. Andersen, M. G. Lewis, D. H. Barouch, DNA vaccine protection against SARS-CoV-2 in rhesus macaques. *Science* 369, 806–811 (2020).
32. J. E. Bryant, A. S. Azman, M. J. Ferrari, B. F. Arnold, M. F. Boni, Y. Boum, K. Hayford, F. J. Luquero, M. J. Mina, I. Rodriguez-Barraquer, J. T. Wu, D. Wade, G. Vernet, D. T. Leung, Serology for SARS-CoV-2: Apprehensions, opportunities, and the path forward. *Sci. Immunol.* 5, eabc6347 (2020).
33. C. Anderson-Bergman, icenReg: Regression Models for Interval Censored Data in R. (2017). <https://www.jstatsoft.org/article/view/v081i12>

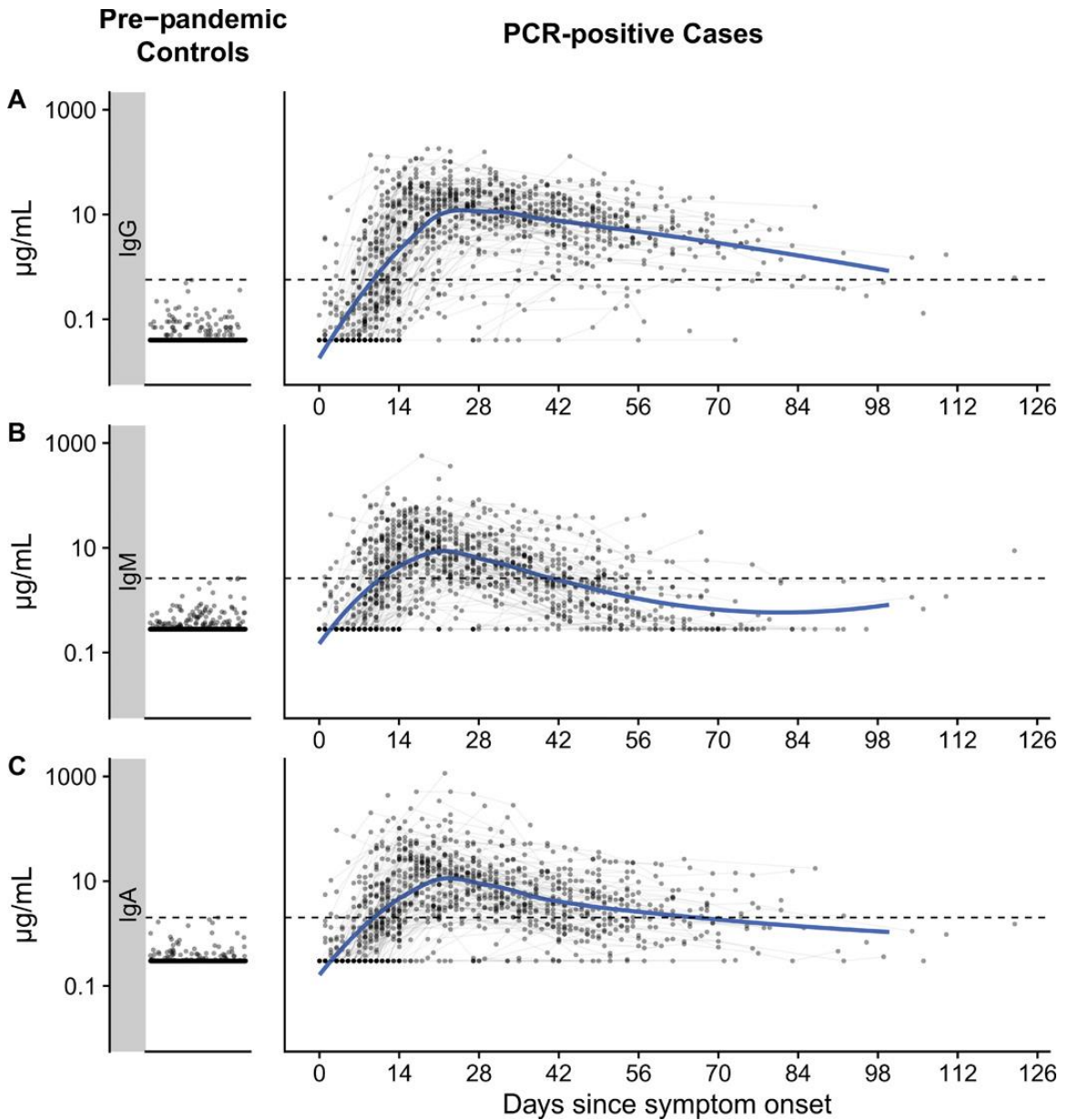
**Table 2.1. Individual characteristics of PCR-positive SARS-CoV-2 cases and pre-pandemic controls.**

\*Pre-pandemic controls included healthy adults (n=274), patients undergoing routine serology testing (n=1241), and patients presenting with other known febrile illnesses (n = 33), including 13 with bacteremia (e.g., *S. aureus*, *S. pneumoniae*, *E. coli*, or *K. pneumoniae* confirmed by standard microbiologic techniques), 4 with babesiosis (confirmed by microscopy and/or PCR), 1 with presumed scrub typhus, and 15 with viral respiratory infections (e.g., influenza [7], parainfluenza [4], respiratory syncytial virus [3], and metapneumovirus [1] confirmed by PCR or direct fluorescent antibody test).‡Data available for 310 cases. †Data available for 342 cases.

Characteristic	Pre-pandemic Controls* (N=1,548)	PCR-positive Cases (N=343)
<b>Age</b>		
Median [IQR]	37 [30–54]	59 [45–71]
<65 years (%)	1,386 (90)	213 (62)
65+ years (%)	162 (10)	130 (38)
Female (%)	1,024 (66)	132 (38)
<b>Race or ethnic group‡</b>		
White (%)	NA	125 (36)
Black or African American (%)	NA	34 (10)
Hispanic or Latino (%)	NA	121 (35)
Asian, American Indian, Alaska Native or Other (%)	NA	30 (9)
Immunosuppressed (%)	NA	26 (8)
<b>Severity†</b>		
Not Hospitalized (%)	NA	24 (7)
Hospitalized, no ICU (%)	NA	138 (40)
Hospitalized, required ICU (%)	NA	137 (40)
Died due to COVID-19 (%)	NA	43 (13)

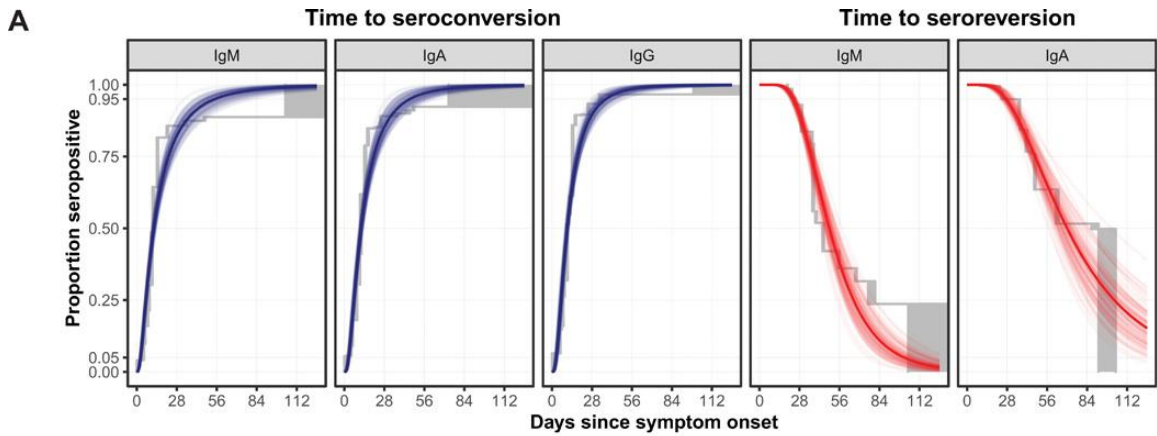
**Figure 2.1. Measurement of IgG, IgM, IgA against SARS-CoV-2 spike protein receptor binding domain among pre-pandemic controls and PCR positive cases.**

Each dot represents a unique measurement of an isotype (Row A: IgG, Row B: IgM, Row C: IgA) in pre-pandemic controls (left panels) and PCR positive cases (right panels). The blue line is a loess smooth nonparametric function. Black dashed lines indicate the maximum concentration ( $\mu\text{g/mL}$ ) found among pre-pandemic controls (IgG: 0.57, IgM: 2.63, IgA: 2.02). Horizontal jitter was introduced into the pre-pandemic controls. The limit of detection ( $\mu\text{g/mL}$ ) was 0.04 for IgG, 0.28 for IgM, and 0.30 for IgA.



**Figure 2.2. Parametric and nonparametric model estimates of time to seroconversion and seroreversion for each isotype.**

A) The isotype cut-offs chosen for seroconversion were the maximum concentration ( $\mu\text{g/mL}$ ) found among pre-pandemic controls (IgG: 0.57, IgM: 2.63, IgA: 2.02). The solid line represents the estimated cumulative distribution function of the time to seroconversion or reversion with 100 bootstrapped fits shown as transparent lines. The parametric accelerated failure time models assume a log-normal time-to-event distribution. Nonparametric estimates shown in grey were calculated using the Turnbull method. Only 3 individuals seroreverted for IgG, so no model is included. B) The table indicates the estimated average number of days since onset of symptoms it takes for a percentage of cases to seroconvert or serorevert. Bootstrap 95% confidence intervals are shown in parentheses.



**B**

Type	Isotype	5th Percentile	25th Percentile	50th Percentile	75th Percentile	95th Percentile
Seroconversion	IgA	3.0 (2.2-3.9)	6.7 (5.7-7.6)	11.7 (10.4-13.0)	20.5 (17.5-24.2)	46.5 (33.9-62.4)
	IgG	3.1 (2.1-4.1)	6.4 (5.3-7.5)	10.7 (9.6-11.9)	18.0 (15.9-20.7)	38.2 (29.6-49.8)
	IgM	2.6 (1.9-3.5)	6.4 (5.4-7.5)	11.9 (10.5-13.4)	22.2 (18.5-26.9)	54.9 (38.8-76.2)
Seroreversion	IgA	27.7 (22.8-32.9)	47.9 (42.1-54.9)	70.5 (58.5-87.5)	104.0 (78.5-142.9)	183.3 (116.8-294.8)
	IgM	23.7 (21.6-26.0)	36.3 (33.5-39.7)	48.9 (43.8-55.6)	65.9 (56.5-78.7)	101.4 (80.8-130.7)

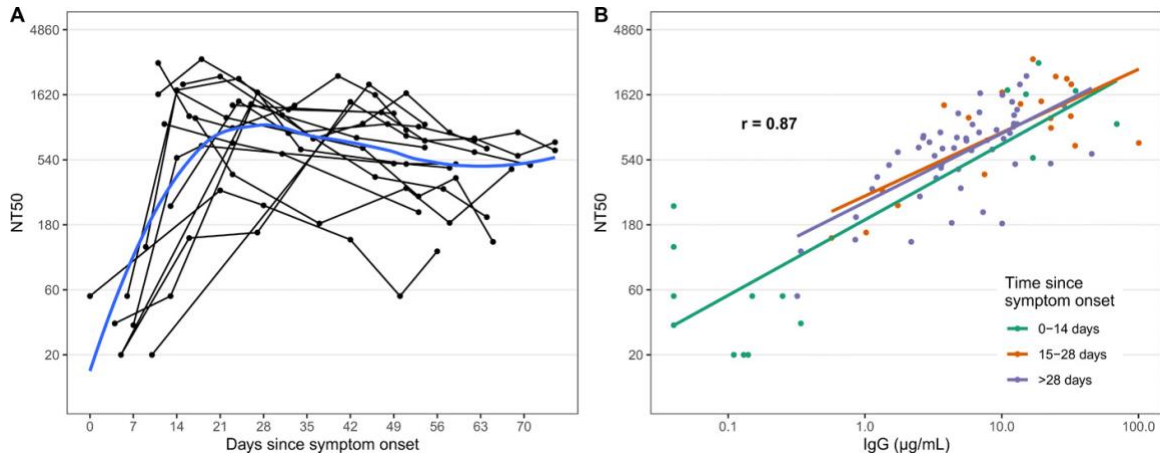
**Table 2.2. Predictive accuracy of individual isotypes for classifying controls and cases across time.**

The isotype cut-offs chosen for calculating sensitivity were the maximum value found among pre-pandemic controls (IgG: 0.57  $\mu\text{g/mL}$ , IgM: 2.63  $\mu\text{g/mL}$ , IgA: 2.02  $\mu\text{g/mL}$ ). Bootstrap 95% confidence intervals are shown in parentheses.

Isotype	Days since symptom onset	AUC (95% CI)	Sensitivity (95% CI)
IgG	$\leq 7$ days	0.68 (0.66–0.70)	0.07 (0.03–0.12)
	8-14 days	0.91 (0.89–0.92)	0.51 (0.43–0.58)
	15-28 days	0.99 (0.99–1.00)	0.95 (0.92–0.98)
	>28 days	0.99 (0.99–1.00)	0.95 (0.91–0.98)
IgA	$\leq 7$ days	0.63 (0.61–0.65)	0.07 (0.03–0.11)
	8-14 days	0.87 (0.85–0.89)	0.44 (0.38–0.51)
	15-28 days	0.98 (0.97–0.98)	0.89 (0.84–0.94)
	>28 days	0.98 (0.97–0.98)	0.60 (0.51–0.68)
IgM	$\leq 7$ days	0.60 (0.58–0.62)	0.08 (0.03–0.13)
	8-14 days	0.87 (0.85–0.89)	0.55 (0.48–0.62)
	15-28 days	0.98 (0.97–0.99)	0.86 (0.81–0.92)
	>28 days	0.93 (0.91–0.94)	0.51 (0.43–0.59)

**Figure 2.3. SARS-CoV-2 pseudovirus neutralization antibody titers in symptomatic PCR positive cases and correlation with anti-RBD IgG responses.**

A) Each point represents a measurement of 50% neutralizing titer (NT50). Lines connect measurements from the same individual and a loess smooth function is shown in blue. B) The overall repeated measures correlation coefficient ( $r$ ) is shown. Lines represent simple linear models for each time period.



## **CHAPTER 3: Identifying recent cholera infections using a multiplex bead serological assay**

Forrest K. Jones, Taufiqur R. Bhuiyan, Rachel Mills, Ashraful I Khan, Damien Slater, Kian Robert Hutt Vater, Fahima Chowdhury, Meagan Kelly, Peng Xu, Pavol Kováč, Rajib Biswas, Mohammad Kamruzzaman, Edward T. Ryan, Stephen B. Calderwood, Regina C. LaRocque, Justin Lessler, Richelle C. Charles, Daniel T. Leung, Firdausi Qadri, Jason B. Harris, Andrew S. Azman

### **Abstract**

Population-level incidence of cholera infections can be estimated from cross-sectional serological data. Current laboratory methods are resource intensive and challenging to standardize across laboratories. A multiplex bead assay (MBA) could efficiently expand the breadth of measured antibody responses and improve accuracy. We tested 305 serum samples from confirmed cholera cases (collected 2-1083 days post-infection) and uninfected household contacts in Bangladesh for serological biomarkers using an MBA (IgG, IgA, and IgM for 7 *Vibrio cholerae* O1-specific antigens related to infection) as well as the vibriocidal assay (Ogawa and Inaba serotypes) and enzyme-linked immunosorbent assay (IgG and IgA for 2 antigens). While post-infection vibriocidal responses often had much higher initial fold-changes (52 and 50 times on-average) than other markers, several MBA-measured antibodies



demonstrated robust responses with similar or longer half-lives. Random forest models combining all MBA antibody and isotype measures allowed for accurate identification of cholera infections: a model identifying infection in the last 200 days had an estimated a cross-validated AUC for ( $cvAUC_{200}$ ) of 92%. Simplified models based on only 3 IgG antibody responses had similar accuracy ( $cvAUC_{200} = 89\%$ ). Across different infection windows (between 45- and 300-days), predictive accuracy of models trained on MBA measurements were non-inferior to models based on traditional serological assays. An MBA-based seroincidence assay can allow for expanded serosurveillance efforts globally including the use of these antigens in multi-pathogen serosurveillance platforms.

## **Introduction**

Cholera remains a global public health threat with an estimated 95,000 deaths per year, especially in areas without access to safe water and adequate sanitation (1). Seventh pandemic strains of *V. cholerae* (toxigenic serogroup O1 El Tor biotype) are responsible for most cholera cases, with endemic transmission in subnational areas of Africa and South Asia as well as large outbreaks in conflict zones, humanitarian crises and post-disaster settings (2–5). Several countries plan to achieve large reductions in cholera cases and deaths over the next decade using a multisectoral approach through the use of oral cholera vaccines and investments in water and sanitation infrastructure (6). A clear understanding of the magnitude of *V. cholerae* transmission at the national and sub-national level is essential for targeting and monitoring of global progress towards ending cholera.

Cholera surveillance typically consists of clinic-based syndromic surveillance for acute watery diarrhea with infrequent laboratory confirmation (7). When laboratory confirmation is performed, often less than half of suspected cholera cases have detectable *V. cholerae* by culture, though this varies considerably by setting (4, 5, 8). Since most infections with *V. cholerae* lead to mild or no symptoms, clinical surveillance detects only a small fraction of infections (9, 10). Clinical surveillance systems are also subject to biases related to individual healthcare access and design of the surveillance system (e.g. sentinel sites) (11, 12). As a result, clinical surveillance alone provides a skewed understanding of disease burden and transmission of *V. cholerae*.

Serosurveillance has been a useful complement to clinical surveillance for a variety of pathogens and there is growing interest in its use for monitoring cholera incidence (13, 14). Despite variability in clinical outcomes, infection with cholera, regardless of symptoms, typically leads to a robust measurable immune response, including a rise, and eventual decay, in serum-circulating antibodies against multiple epitopes (15). As a result, cross-sectional measurements of circulating antibodies can provide insights into the incidence and timing of past infections. The two most common methodologies used to measure antibodies generated in response to *V. cholerae* infection are the vibriocidal assay and enzyme linked immunosorbent assay (ELISA). Previous studies have shown that vibriocidal titers rise quickly after infection and then decay toward pre-infection levels after 1 year (15, 16). However, the vibriocidal method is a functional assay that requires culturing *V. cholerae* over several hours (thus requiring Biosafety

Level 2 practices, equipment, and facility design) and is challenging to standardize across laboratories (17). Although ELISA tests targeting IgG and IgA antibodies that bind to known antigens are easier to implement, these assays are less predictive of cholera infection than the vibriocidal assay (18). Recent work illustrates that combining vibriocidal titers with ELISA antibody measurements in statistical models can identify individuals infected in last year for the purposes of estimating cholera seroincidence (i.e., the incidence of meaningful immunologic exposures to *V. cholerae* O1 over a specific time period) (18).

Over the past decade, advances in high-throughput multiplex bead serological assays (MBA) have enabled their use to study the burden, risk, and dynamics of a variety of pathogens (19–22). These assays only require a small volume of serum (e.g., 1  $\mu$ L to measure multiple antigens [when performed in duplicate] as compared to 12.5  $\mu$ L for the vibriocidal assay), potentially are more sensitive (23), and could be easier to standardize (24). Additionally, they allow for the characterization of multiple antigens simultaneously, improving the efficiency and cost of the assay as compared to running multiple ELISAs (25). This also facilitates broad exploration of novel antigens that may correlate with previous exposure or immunity. If measuring multiple antibodies to *V. cholerae* antigens is as predictive of previous infection as the vibriocidal assay, serosurveillance would be more feasible in many more settings. However, the use of cholera antigens in an MBA to predict previous infection has not been previously assessed.

Here, we characterize post-infection antibody dynamics to seven cholera antigens up to three years post-infection in a cohort of confirmed medically-attended *V. cholerae* O1 infections. We use these serological data to fit machine learning models aimed at identifying recently infected individuals. We then compare performance of models based on this assay to those based on traditional antibody measurements and suggest a reduced panel of antigens to be used in MBA arrays for future cholera serosurveillance efforts.

## **Methods**

### Study Population

As described previously, consenting patients hospitalized at the icddr,b (formerly known as International Centre for Diarrhoeal Disease Research, Bangladesh) Dhaka hospital with culture-confirmed *V. cholerae* O1 infection were enrolled between 2006 and 2018 (26, 27). These cases were followed up to 1083 days (~3 years) post enrollment with blood samples collected periodically. We approximated the number of days between infection and sample collection by taking the difference between the enrollment date and sample collection date, then adding the number of hours of symptomatic diarrhea prior to enrollment (range: 3-60 hours), then adding 1.4 days for the incubation period (28), and finally rounding to a whole number of days. Household contacts of confirmed cases (defined as individuals who shared a cooking pot with the index case for three or more days preceding the cholera episode in the index case) were also enrolled with blood and stool samples collected at approximately 2-, 7-, and 30-days post enrollment of the initial cases. We limited our selection of household

contacts to those that had no evidence of cultured *V. cholerae* from stool samples during follow-up. Data on non-immunological variables of age, sex, and blood type were available for all participants.

Prior to this study, serum samples of cases and contacts had been tested using vibriocidal assays (Ogawa and Inaba serotypes) and ELISAs (26, 27). The latter included measures of both IgG and IgA for lipopolysaccharide Ogawa serotype (LPS) and cholera toxin B subunit (CT-B) antigens (18). To conserve limited specimens and lab reagents, we selected 20 individuals <10 years old and 20 individuals  $\geq 10$  years old to have a balance of cases for whom their infection was either their first or a subsequent infection. To further ensure a balanced distribution of different serological trajectories, we selected these 40 individuals such that when fit with a model, led to the most accurate classification of samples from the rest of the cohort (Figure B.1). Specifically, we trained random forest models on 10,000 different potential training sets of 40 individuals and then selected the set that had the highest cross-validated area under the curve (cvAUC) when predicting infection within 1-year for samples in the test set (Supplementary Methods in Appendix B).

After investigation of freezer stocks, samples from 39 of 40 individuals were available for testing. Samples from two additional cases were added to the set, giving a selection of 245 samples from 41 individuals (39 cases and 2 uninfected household contacts). To include more individuals between the age of 10 and 18, we supplemented this initial sample with all samples available from 9 cases and 1 contact from this age group (Figure B.2) (who were randomly

selected from available sera). To limit the influence of boosted antibody responses from reinfection/exposure during the follow-up period, we removed four data points (from two cases) that were part of or after a greater than 2-fold-rise between measurements in vibriocidal Ogawa titers >90 days post initial infection. As all rises greater than 2-fold for vibriocidal Inaba titers followed a previous drop in titers below a detectable range (which were considered unlikely to be a part of reinfection pattern), only the Ogawa serotype was used for identifying reinfections.

#### *Serological testing and data processing*

Based on previous work on the immune responses to *V. cholerae* infection (29–31), we selected a panel of eleven antigens to investigate with a multiplex bead assay. Seven of these had been previously shown to be stimulated by infection with *V. cholerae* O1. These included O1 serogroup Ogawa serotype O-specific polysaccharide (OSP, part of the LPS), O1 serogroup Inaba serotype OSP, CT-B, cholera toxin holotoxin (CT-H), Toxin co-regulated pilus subunit A (TcpA), *V. cholerae* cytolysin (VCC) (also known as hemolysin A), and *V. cholerae* sialidase. We also investigated the serum antibody response against four control antigens: O139 serogroup OSP (*V. cholerae* O139 has rarely been detected in humans over the last decade, but is included in most oral cholera vaccines), heat labile enterotoxin subunit B (LT-B), and heat labile holo-enterotoxin (LT-H) (expressed during infection with enterotoxigenic *Escherichia coli* [ETEC] with a high degree of homology with cholera toxin) were also selected. Influenza haemagglutinin 1 (Flu) was also included as a control antigen.

All antigens were conjugated to magnetic beads using the procedures specified by Luminex (32).

Each plate included a dilution series (from pooled convalescent sera of 5 patients with culture-confirmed *V. cholerae* O1 infection) and control wells, all of which were run in triplicate. Following the testing protocol, serum, beads, and secondary antibodies binding to IgG, IgA, and IgM were added to each well. Samples were run on a Luminex Flexmap 3D machine at Massachusetts General Hospital by one technician. Bead counts and median fluorescence intensity (MFI) values were exported from the Exponent software program. Plates were retested when over half of the positive control dilutions had  $\geq 5$  antigens with a coefficient of variation (calculated from triplicate MFI measurements) greater than 20%.

For the analysis, any measurements with a bead count less than 30 were excluded ( $< 0.1\%$ ). MFI values were averaged across replicate wells. We standardized MFI values from the assay to help adjust for inter-plate variability by calculating the relative antibody unit (RAU) (Supplementary Methods in Appendix B) (33). For each plate, we fit a 4-parameter log-logistic model to the dilution series and used the median of parameter estimates to predict the RAU for each sample. (Figure B.3, B.4, B.5, and B.6) (34, 35). For samples with a predicted RAU outside the range of  $10^5$  and  $10^2$ , the RAU was set at the threshold value. Additionally, we calculated the Net MFI for each sample (i.e., MFI of sample - MFI of blank well, but censored at 10 FI units). Despite some between-plate variability and limits of detection, we observed high correlation between the Net MFI and RAU measurements, across time points.

## Statistical Analysis

We fit hierarchical regression models for each marker to estimate the degree of antibody boosting post-infection and its decay rate after the boost for each serological marker. We used a Bayesian framework with two components: a kinetic model and a measurement model (Supplementary Methods in Appendix B, (36)). For the kinetic model, we assumed individuals had an instantaneous boost of antibodies 5 days post-infection followed by decay over time and explored alternate lags in sensitivity analyses. We explored both exponential and biphasic models for antibody decay and compared the expected predictive log density values between both parameterizations. In the measurement model, we assumed random error was normally distributed (on the log-scale) and accounted for the fact that some observations were censored (e.g. titration data). We fit the models using Markov Chain Monte Carlo methods implemented in Stan (35, 37). We also investigated how age group (<10 years vs. ≥10 years), sex (male vs. female), blood type (O blood type vs. non-O blood type), and infecting serotype (Ogawa vs. Inaba) affected baseline antibody levels, boosting, and decay by including fixed effects for these binary variables in the model (Supplementary methods in Appendix B).

We also explored the ability of statistical models to identify individuals who were recently infected with *V. cholerae* O1. Using serological biomarkers and three non-immunological demographic variables (age, sex, and blood type), we trained random forest classification models to identify recently infected individuals (38). We fit models using several definitions of recent infection (i.e.



the infection window), including having been infected within 45-, 120-, 200-, or 300 days before blood collection. For example, the proportion of individuals classified as recently infected using a 120-day model is what is referred to as 120-day seroincidence. Uninfected household contacts and cases infected 5 or fewer days prior (due to insufficient time to generate an immune response) were always considered as not recently infected. We fit models with weights to account for both class imbalance and for the large concentration of measurements collected during the early convalescent period (7-30 days) compared to later post-infection period (Figure B.7, Supplementary Methods in Appendix B). Using 10-fold cross-validation on the full dataset, we estimated the cvAUC to evaluate the ability of the model to identify recently infected individuals. For each individual, we kept all measurements within the same fold. To understand which markers had the largest influence on model fits, we used a permutation importance metric (39). We fit three alternative models (i.e. Lasso and Elastic-Net Regularized Generalized Linear Models, Bayesian Additive Regression Trees, and Extreme Gradient Boosting) to understand the influence of model choice on our results. We also combined results to yield ensemble predictions and estimate their cvAUC (40).

We also evaluated the specificity and time-varying sensitivity of the random forest classification models using leave-one-individual-out cross-validation. For each fold (i.e. left-out individual), we fit random forest models to 100 random samples of 50% individuals in the training set and found a cut-off that satisfies the Youden Index or a desired specificity cut-off (90%, 95%, or

99%) in the other 50% of individuals. Using the median value of these cut-offs and a model fit with the entire training data, we predicted the serostatus of the left-out samples. Using the predicted serostatus for each sample from leave-one-individual-out cross-validation, we fit hierarchical logistic regression models to estimate the specificity and time-varying sensitivity of each random forest model (10, 41). For time-varying sensitivity models, we assumed that the  $\text{logit}(\text{sensitivity})$  was a function of (log-transformed) days since infection. We allowed for increasingly complex functions as time since infection increased, including a constant sensitivity for the 45-day model, a linear decrease in sensitivity for the 120-day model, a quadratic polynomial for the 200-day model and a cubic polynomial for the 300-day infection window model.

Data and code used to select samples and conduct primary analyses are available at: <https://github.com/HopkinsIDD/cholera-multiplex-panel>.

## **Results**

### *Description of individuals and timing of samples*

We tested 296 samples from 48 confirmed cholera cases (2 to 1,083 days post infection) and 9 samples from 3 uninfected household contacts of cases (Table B.1). *V. cholerae* serogroup O1 was isolated from each case with most being serotype Ogawa (81%) and the rest being serotype Inaba. The median age of cases at time of enrollment was 11 years (interquartile range (IQR): 6-26 years) with 17% being <5 years old and 35% being  $\geq 18$  years old. Most cases were male (62%) and nearly half had the O blood type (46%).

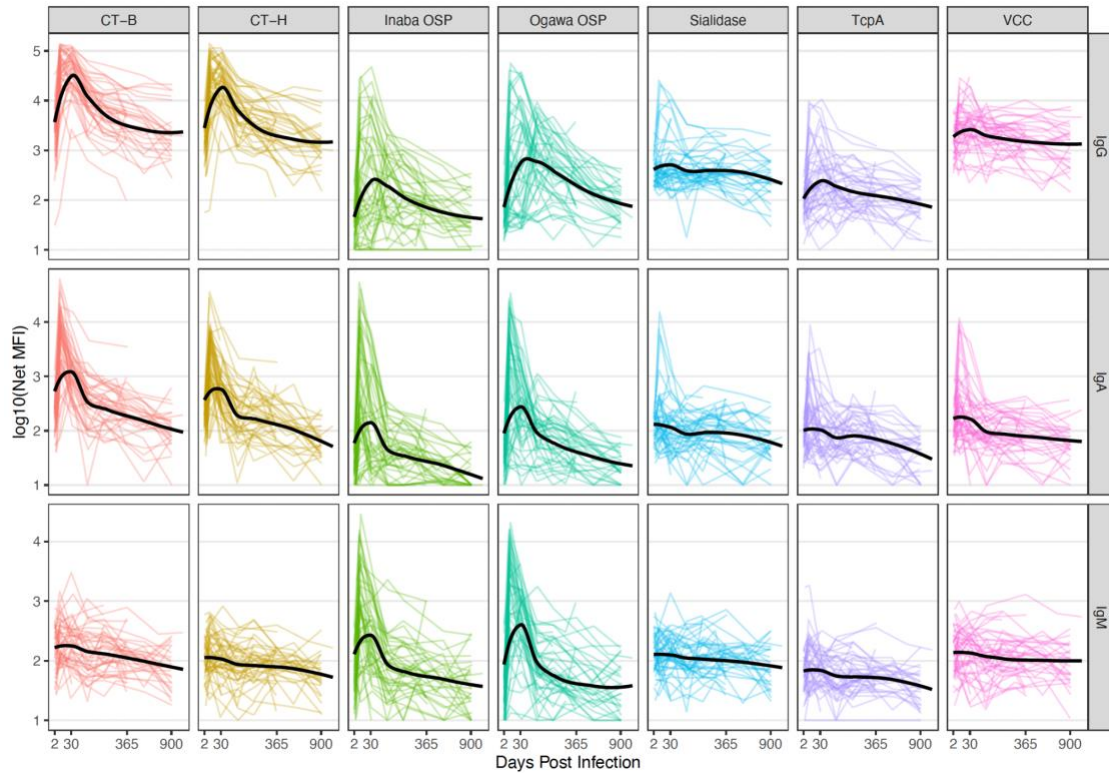
All cases had a baseline sample collected between 2 and 4 days after infection. Nearly all cases had additional samples collected between 7-9 days (n=46), 27-36 days (n=46), 87-109 days (n=42) and 173-191 days (n=39) post symptom onset. Between 269 and 1083 days after infection, 1 person had three samples, 35 people had two samples, 2 people had one sample, and 10 people had zero samples collected (Figure B.2).

### *Kinetics of biomarkers in confirmed cholera cases*

After infection, the levels of several *V. cholerae* O1-specific antibodies among many cases had a steep rise followed by variable decays (Figure 3.1, B.8, B.9, and B.10). Robust anti-CT-B, anti-CT-H, anti-Inaba OSP, and anti-Ogawa OSP antibody responses were observed across the study cohort (except anti CT-B and anti-CT-H IgM potentially indicating an anamnestic response) . Individuals that did have an increase in anti-sialidase, anti-TcpA, and anti-VCC antibodies tended to be adults (Figure B.11). Among anti-CT-B, anti-CT-H, anti-Inaba OSP and anti-Ogawa OSP antibodies, the observed median day of peak measurement was 8 days for IgA, between 8 and 64 days for IgM, and 25 to 34 days for IgG. There were no substantial increases in anti-O139 OSP or anti-Flu antibodies after infection (Figure B.8, Figure B.10), as expected. Some antibody responses were highly correlated due to antigen homology, including cross-reactivity between the similar Ogawa and Inaba OSP antigens, and the CT and LT antigens (30) (Figure B.13).

We fit a series of statistical models to estimate the population-level rise and decay of each marker. As biphasic models did not consistently fit better than

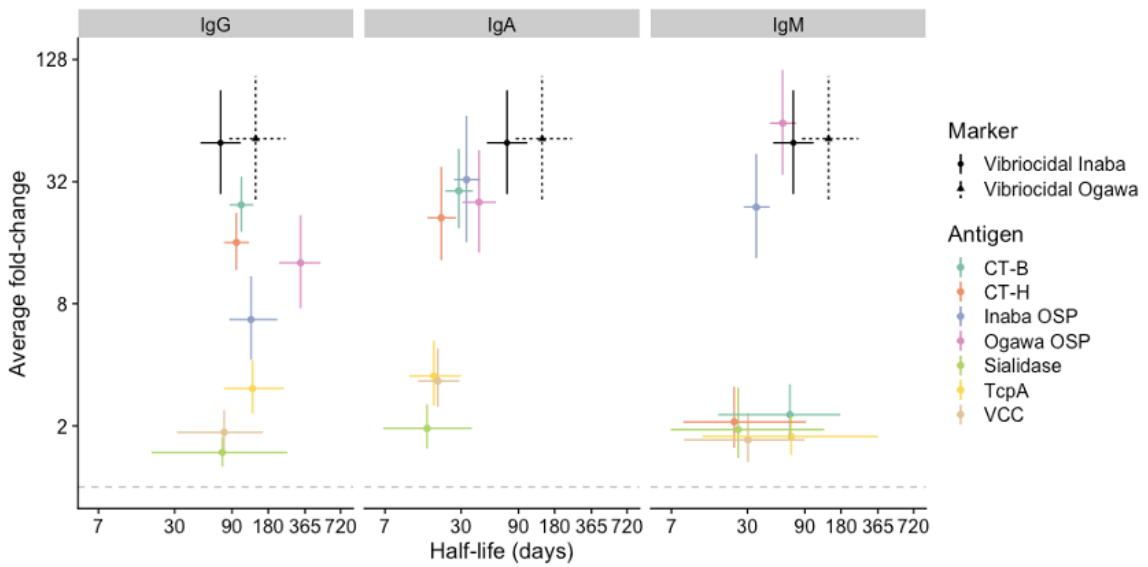
exponential models (i.e., biphasic models had a significantly higher expected predictive log density value for only 10 of 39 markers [Table B.2]), we utilized exponential antibody decay for our analysis. These models were able to reproduce individual-level antibody trajectories (Figure B.14-B.19) reasonably well. The magnitude of initial antibody rise and duration of half-life varied considerably across MBA measures of antibody levels and vibriocidal titers (Figure 3.2, Table B.3). The initial boost in vibriocidal titers (52 average fold-rise for Ogawa and 50 for Inaba) was higher than for all MBA-measured antibodies except anti-Ogawa OSP IgM. We estimated relatively large boosts in anti-Ogawa OSP and anti-Inaba OSP antibodies across isotypes (range: 7 - 62 average fold-rise) and large boosts in anti-CT-B and anti-CT-H IgG and IgA antibodies (range: 16 - 29 average fold-rise). Anti-Ogawa OSP IgG antibodies had the longest estimated half-life (335 days [95% CI: 221-490]). Anti-CT-B IgG, anti-CT-H IgG, and anti-Inaba OSP IgG antibodies had similar half-lives (98-130 days) as the vibriocidal markers (72 and 142 days) while the half-life for IgA and IgM antibodies was generally shorter (range: 11- 69 days). Antibodies measured by ELISA had less pronounced (average fold-rises all less than 2) and short lived (all half-lives less than 76 days) (Table B.3).



**Figure 3.1. Multiplex bead assay measurements of IgG, IgA, and IgM against *V. cholerae* O1 antigens among culture confirmed cholera patients.**

The y-axis indicates the  $\log_{10}(\text{Net MFI})$  and the x-axis is the number of days post-infection (square-root transformed). Each colored line indicates individual trajectories over time. The Black solid line is a loess smooth function. A similar plot with relative antibody units (RAU) can be found in Figure B.8. A plot of other measured antigens can be found in Figure B.9 and Figure B.10.

Both age and infecting serotype influenced antibody kinetics (Figure B.11, Figure B.12). We found that individuals <10-years old tended to have smaller initial anti-Ogawa OSP and anti-Inaba OSP IgG responses (boosts were 3.9 (95% CI 1.4-12.5) and 5.0 (95% CI 1.9-15.8) times smaller) but with slower decay (difference in half-life was 645 days (95% CI 262-1857) and 333 days (95% CI -77-945)) compared to those  $\geq$ 10-years old (Table B.4). Individuals <10-years old had lower baseline and initial responses for anti-Ogawa and anti-Inaba OSP IgA, but the rates of decay were similar. Individuals <10-years old had little difference as compared to those  $\geq$ 10-years old in their anti-OSP IgM and anti-CT-B trajectories for any isotype. Individuals with Ogawa infections on-average had 4.0 (95% CI 1.2-10.9) times higher initial boosts in anti-Ogawa OSP IgM than those with Inaba infections, but similar initial boosts in anti-Ogawa OSP IgG and IgA (Table B.4). Individuals with Inaba infections on-average had higher initial boosts in anti-Inaba OSP IgG (3.3 times, 95% CI: 1.4-9.8) than those with Ogawa infections. There were limited differences in terms of baseline values, boost, or decay rates across individuals with different blood type and sex (Table B.4).



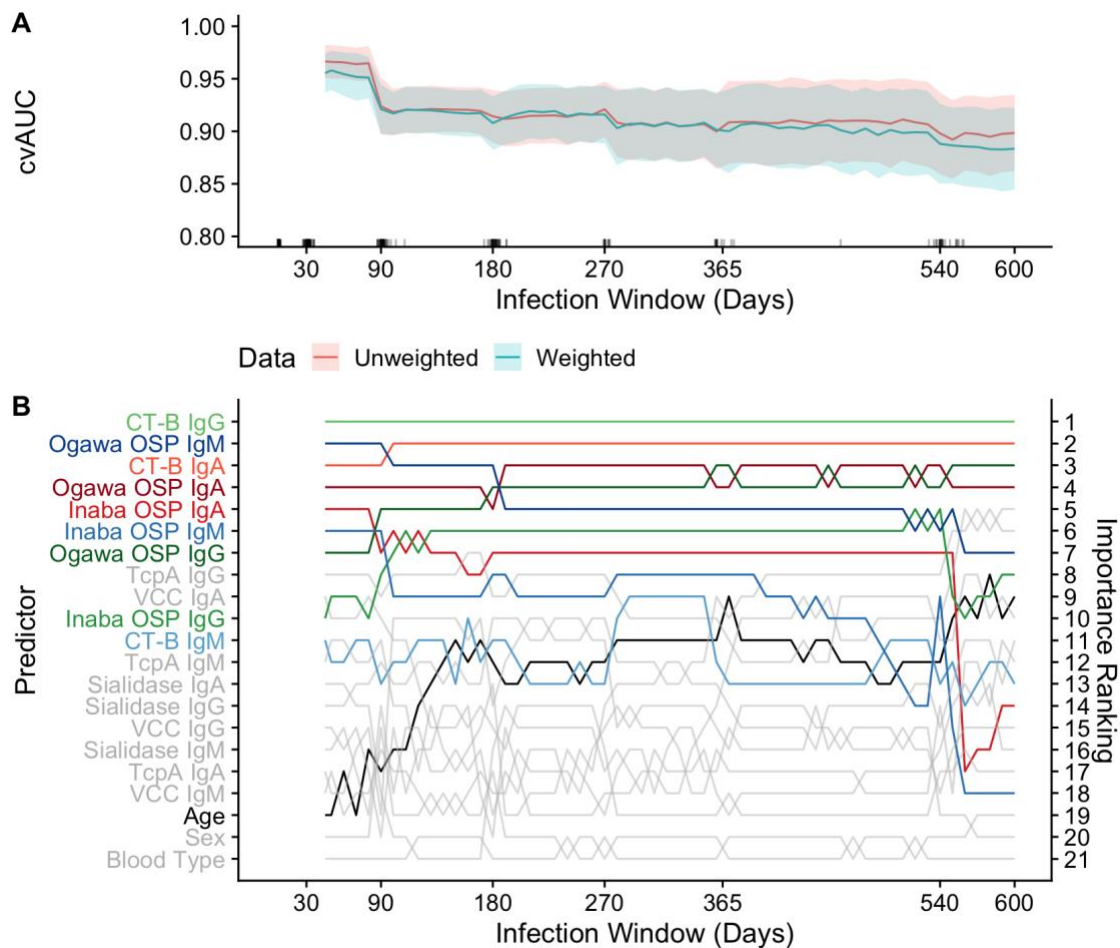
**Figure 3.2. Estimated duration of half-life and average fold-change from exponential decay models.**

Each point indicates the median estimate of the average individual fold-rise from baseline to peak (y-value) and the median estimate of the half-life (x-value) for exponential decay univariate models. Marginal 95% credible intervals are shown as lines. Model estimates for the vibriocidal assays in each panel are identical.

### Identification of recent infections with cross-sectional serologic measurements

We estimated the cvAUC of random forest models trained on 18 MBA markers (six antigens [CT-H was not included given its high correlation with CT-B] and three isotypes) and three individual non-immunological factors to understand their ability to identify recently infected individuals. The average cvAUC was consistently above 88% regardless of infection window but was higher at shorter time windows (Figure 3.3A, Figure B.20). Models using observations that were weighted based on time since infection either performed equally or slightly worse than models without weights (Figure 3.3A). As an ensemble of four different machine learning models performed similarly to the random forest model, we conducted further analyses with the random forest model alone (Figure B.21).





**Figure 3.3. Cross-validated area under the receiver operating characteristic curve (cvAUC) and predictor importance rankings for random forest models trained on MBA markers across infection windows.**

Estimates of mean cvAUC (10-fold) and 95% confidence interval are shown for weighted and non-weighted models between 50- and 600-day infection windows at 10-day intervals (A). Rug plot shows the day of collection of samples from cases used in training models. Samples collected under 5 days since infection, over 600 days since infection, or from household contacts are not shown. For each infection window of weighted models, the rankings of predictors by their importance are shown on the y-axis (B). Colors of lines are unique to each predictor.

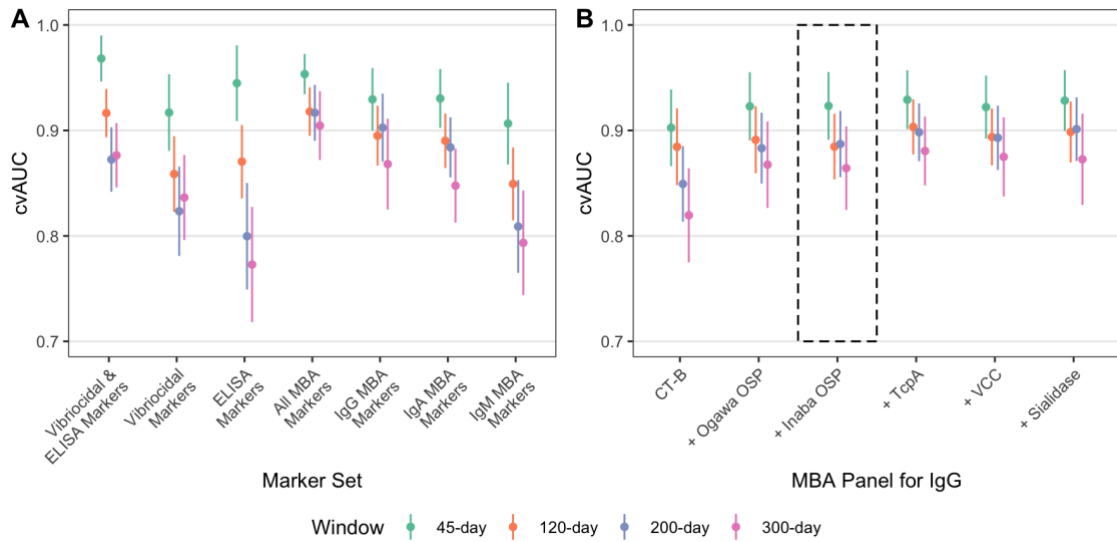
Across infection windows, anti-CT-B IgG antibodies were consistently the most influential predictor of recent infection while the relative importance of predictors changed with different infection windows (Figure 3.3B, Figure B.20). With infection windows shorter than 90 days, anti-Ogawa OSP IgM was the second most influential marker, but waned in influence over longer windows, while the relative influence of anti-Ogawa OSP IgG increased over time. Anti-CT-B IgA and anti-Ogawa OSP IgA were consistently among the most influential markers. Anti-Ogawa OSP markers were always more influential than anti-Inaba OSP markers within isotype (likely because 81% of cases used to train the models were infected with the *V. cholerae* Ogawa serotype). Other antibodies, age, sex, and blood type did not greatly influence the model.

We then compared cvAUC from models fit to MBA measurements with those fit with the traditional vibriocidal and ELISA measurements for four infection windows (45-day, 120-day, 200-day and 300-day). We removed 11 (4%) of samples (from cases) for this analysis as they were either missing a vibriocidal or ELISA measurement. The model fit with both vibriocidal and ELISA markers was highly predictive of recent infection at 45-day (cvAUC: 97%), 120-day (cvAUC: 92%), 200-day (cvAUC: 87%) and 300-day (cvAUC: 88%) infection windows (Figure 3.4A, Table B.5). The cvAUC of models trained with all 18 MBA markers was consistently similar to models trained with vibriocidal and ELISA markers (range of cvAUC ratios: 0.98-1.05). (Figure 3.4A, Table B.5)

### Multiplex bead assay panel simplification

Reducing the number of isotypes tested may be necessary for the integration of cholera serosurveillance into larger MBA-based antigen panels. We explored how using fewer MBA markers would impact model performance (Figure 3.4A, Table B.5). Across all timescales, a model using all 6 IgG MBA markers (e.g., 200-day cvAUC: 90%, 2% lower mean cvAUC) was slightly more predictive than those using all 6 IgA MBA markers (e.g., 200-day cvAUC: 89%, 4% lower mean cvAUC) or all 6 IgM MBA markers (e.g., 200-day cvAUC: 81%, 12% lower mean cvAUC).

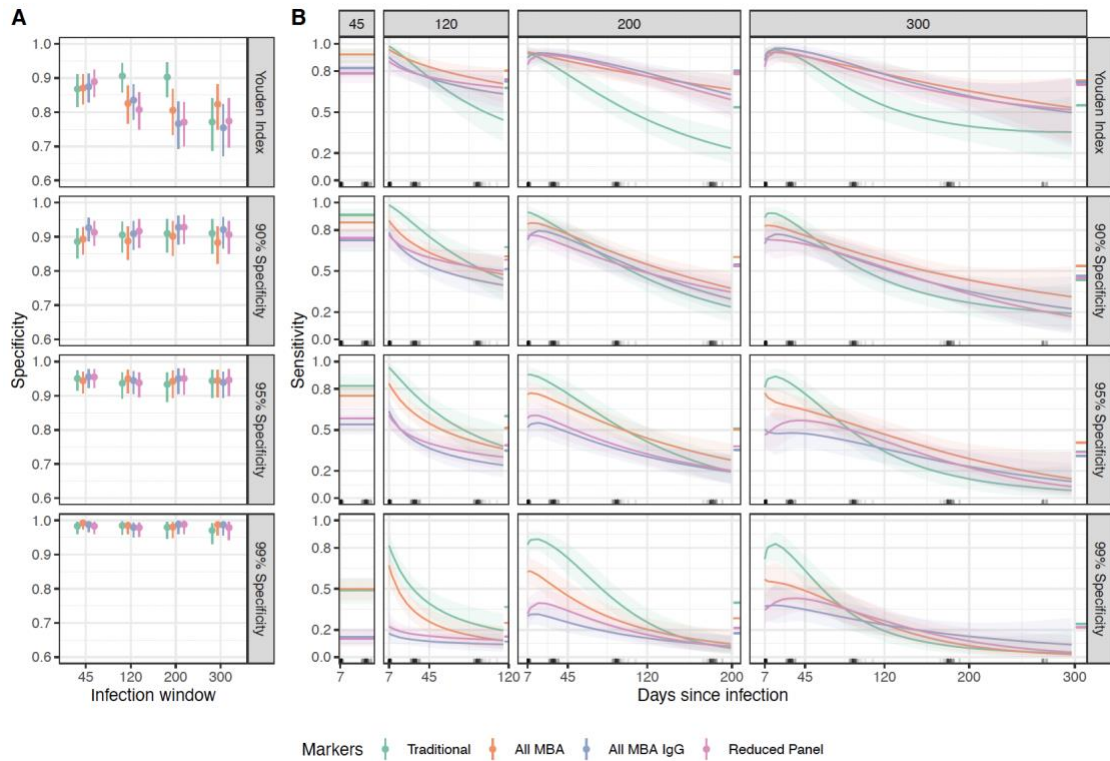
As many commonly used serosurveillance panels are based on IgG markers alone, (42) we considered a reduced panel with only IgG (Figure 3.4B, Table B.6, Figure B.22). A model using only anti-CT-B IgG and individual predictors was predictive (200-day cvAUC: 85%) of cholera infection across all infection windows. Adding both Ogawa OSP and Inaba OSP led to additional improvement (200-day cvAUC: 89%). The addition of TcpA offered a minimal improvement in cvAUC (200-day model: 90% [95% CI: 87%-93%]). The addition of VCC and Sialidase had little impact on the overall cvAUC. We saw similar model performance without inclusion of age, sex, and blood type in these models.



**Figure 3.4. Comparison of cross-validated AUC across random forest models trained on traditional and MBA serological markers for 45-day, 120-day, 200-day, and 300-day infection windows.**

Random forest models were fit using a specified marker set and individual level factors including age, sex, and blood type (A). Estimated mean and 95% confidence intervals for cvAUC are reported. Models fit to reduced panels of IgG MBA markers are shown (B). The order of how antigens were added was determined by the variable importance when fitting a model with only IgG MBA markers. The black dashed box indicates the suggested minimum antigen panel for detecting recent cholera infection.

We then estimated the specificity and time-varying sensitivity of random forest models fit with traditional and MBA markers using leave-one-out cross-validation (Figure 3.5). When using the Youden Index (i.e., jointly maximizing sensitivity and specificity), median estimates of specificity were all below 90% (Figure 3.5A). In general, sensitivity estimates were negatively correlated with higher specificity and did not vary substantially between models trained on different marker sets when using the same infection window (Figure 3.5B). When fixing specificity at 90%, the (non-time-varying) median sensitivity estimates ranged from 72% to 91% for the 45-day infection window. As for other infection windows, sensitivity steadily decreased over time for the 120-day (range of median estimates: 39%-98%), 200-day (range of median estimates: 24%-93%), and 300-day (range of median estimates: 16%-92%) window.



**Figure 3.5. Specificity and time-varying sensitivity estimates of random forest models trained with leave-one-out cross-validation for 45-day, 120-day, 200-day, and 300-day infection windows using different cut-offs.**

Median and 95% credible intervals are shown for the estimated (A) nominal specificity and (B) time-varying sensitivity. Each row represents a different method for acquiring a cut-off including the Youden Index or maximizing sensitivity for a desired value of specificity. The relationship between  $\text{logit}(\text{sensitivity})$  and time since infection (log-transformed) was constant for the 45-day window, linear for the 120-day, quadratic for the 200-day window, and cubic for the 300-day window. Traditional = Vibriocidal Ogawa, Vibriocidal Inaba, and 4 ELISA markers, All MBA = 18 MBA markers, All MBA IgG = 6 MBA markers, Reduced panel = anti-Ogawa OSP, anti-Inaba OSP, and anti-CT-B IgG. All models also included age, sex, and blood type as predictors.

## Discussion

With the aim of increasing the number of tools for serosurveillance of cholera, we developed an MBA, characterized post-infection antibody kinetics in sera of confirmed cholera cases, and estimated how well predictive models trained on these measurements identified recently infected individuals. After infection, anti-OSP and anti-CT-B antibody boosts were the highest and had the longest half-lives, similar to or greater than traditional vibriocidal titers. Models using MBA-measured antibody responses could identify individuals infected up to 300 days before blood collection with similar performance as models employing both the vibriocidal assay and ELISA methods. Simplified models based on as few as three antibody targets performed well and may allow for an expansion of cholera serosurveillance efforts.

The dynamics of the vibriocidal assay and some MBA marker measurements made them valuable for detecting exposure to cholera at different infection windows less than a year. In comparison to MBA markers, ELISA markers had lower post-infection boosts and shorter half-lives, thereby lowering their sensitivity over longer infection windows and limiting their use. Combining the predictive power of MBA markers was generally non-inferior to that of traditional markers for detecting recent infection. However, traditional markers were slightly more sensitive when using a shorter infection window or early-on in longer infection windows than MBA markers.

While we measured robust rises in anti-OSP and anti-CT antibodies among most cases, the comparatively small average rise in antibodies against

VCC, sialidase, and TcpA illustrated that these additional antigens likely have less utility for estimating seroincidence. This may have been partly because our sample had a disproportionately large number of children who were less likely to generate a response to these antigens (Figure B.11). In a previous study of challenged adults, repeated exposures to *V. cholerae* were required to consistently produce antibodies to TcpA (43). We recommend that any MBA panel used to estimate cholera seroincidence from serological surveys at least measure anti-CT-B, anti-Ogawa OSP, and anti-Inaba OSP antibodies. Measuring antibodies for both serotypes might improve a panel's versatility given the distribution of Inaba and Ogawa serotypes can vary considerably across space and time (44).

With growing use of multi-pathogen integrated serosurveillance using MBAs, our study illustrates how inclusion of three additional beads (CT-B, Ogawa OSP, and Inaba OSP) to larger panels could efficiently provide data on cholera incidence. While many serosurveillance panels only measure IgG antibodies (42), technological advancements to simultaneously measure multiple isotypes could further improve seroincidence estimates of cholera infections (45). Standardization of MBA measurements between laboratories will be key to utilizing cut-offs (Figure B.23) or statistical models to classify individuals. This includes using the same antigens, common positive control sera, standardized bead conjugation techniques, and harmonized protocols for running the assay and monitoring quality.



External validation of our findings using serum from individuals infected with *V. cholerae* across the spectrum of clinical severity, more granular age groups, and epidemiologic settings is needed. Our study population consisted of hospitalized cholera cases, which likely have more robust immune responses to cholera (9). Therefore, we might expect statistical models from our study to misclassify mild and asymptomatic infections as not recently infected. Additionally, all members of our study population were from Bangladesh where cholera is endemic, and many individuals are likely infected several times over their lifetime (10). We expect validation of the MBA in an epidemic setting would yield similar results that have been shown with traditional serological assays in challenged North American volunteers (18).

Further investigation of cross-reactive immune responses from cholera vaccination or ETEC infection is required to understand their potential for causing misclassification as recently infected with *V. cholerae*. Specifically, antibodies generated during an immune response to ETEC might bind to CT-B while oral cholera vaccination might stimulate antibodies that bind to OSP. Serosurveys in populations where there is a potential for either vaccination or ETEC to generate cross-reactive immune responses will require additional strategy to limit misclassification. Incorporation of additional antigens into MBA panels or measuring multiple isotypes (Figure B.22) might provide additional serological data that reduce prediction error. For example, most cholera vaccines include an uncommon serogroup of *V. cholerae* (i.e. serogroup O139); antibodies binding to O139 OSP could be considered a signal of recent vaccination. Strategies to

restrict serosurveys to populations or time periods such that the study population likely has few cross-reactive antibodies might also be considered to tackle this problem.

As large investments in cholera prevention and control measures are being made, serosurveillance may be an important tool for tracking trends in incidence to better target interventions and measure their effectiveness in reducing infections. We show how measuring responses to as few as three antibodies with MBA can help identify individuals infected up to one year before, with similar precision as traditional methods. While cholera specific panels may be warranted in some locations, inclusion of *V. cholerae* specific beads in larger multi-pathogen MBAs being used in areas with cholera could lead to the efficient generation of data to aid the fight against cholera.

## **Acknowledgments**

We thank Stephen Lauer for his initial analyses in the selection of serum samples for MBA testing and the Infectious Disease Dynamics Group at Johns Hopkins University for insight on the statistical analysis. We are grateful to Dr. Slavomír Bystrický· Institute of Chemistry, Slovak Academy of Sciences, Bratislava, Slovak Republic for the provision of *V. cholerae* O139 OSP. We acknowledge the study participants and families who consented to enroll into this study. *Funding:* This research was supported through programs funded by the National Institutes of Health, including the National Institute of Allergy and Infectious Diseases including R01 AI137164 (JBH, RCC), R01 AI106878 (ETR, FQ), U01 AI058935, U01 HD39165 (SBC, FQ, ETR), R01 AI135115 (DTL, ASA),

the Fogarty International Center, Training Grant in Vaccine Development and Public Health (TW005572 [RB, MK]), and Emerging Global Fellowship Award TW010362 (TRB), and the Intramural Research Program of the NIH and NIDDK (PX and PK). We are grateful to the Governments of Bangladesh, Canada, Sweden and the UK for providing core/unrestricted support to icddr,b.

## References

1. Roadmap 2030, (available at <https://www.gtfcc.org/about-gtfcc/roadmap-2030/>).
2. J. Lessler, S. M. Moore, F. J. Luquero, H. S. McKay, R. Grais, M. Henkens, M. Mengel, J. Dunoyer, M. M'bangombe, E. C. Lee, M. H. Djingarey, B. Sudre, D. Bompangue, R. S. M. Fraser, A. Abubakar, W. Perea, D. Legros, A. S. Azman, Mapping the burden of cholera in sub-Saharan Africa and implications for control: an analysis of data across geographical scales. *Lancet*. **391**, 1908–1915 (2018).
3. M. Ali, A. R. Nelson, A. L. Lopez, D. A. Sack, Updated global burden of cholera in endemic countries. *PLoS Negl. Trop. Dis.* **9**, e0003832 (2015).
4. A. Camacho, M. Bouhenia, R. Alyusfi, A. Alkohlani, M. A. M. Naji, X. de Radiguès, A. M. Abubakar, A. Almoalmi, C. Seguin, M. J. Sagrado, M. Poncin, M. McRae, M. Musoke, A. Rakesh, K. Porten, C. Haskew, K. E. Atkins, R. M. Eggo, A. S. Azman, M. Broekhuijsen, M. A. Saatcioglu, L. Pezzoli, M.-L. Quilici, A. R. Al-Mesbahy, N. Zagaria, F. J. Luquero, Cholera epidemic in Yemen,

2016–18: an analysis of surveillance data. *The Lancet Global Health*. **6**, e680–e690 (2018).

5. F. K. Jones, J. F. Wamala, J. Rumunu, P. N. Mawien, M. T. Kol, S. Wohl, L. Deng, L. Pezzoli, L. H. Omar, J. Lessler, M.-L. Quilici, F. J. Luquero, A. S. Azman, Successive epidemic waves of cholera in South Sudan between 2014 and 2017: a descriptive epidemiological study. *Lancet Planet Health*. **4**, e577–e587 (2020).

6. WHO | Ending Cholera (2017) (available at <http://www.who.int/cholera/publications/global-roadmap/en/>).

7. G. T. F. on Cholera Control (Surveillance Working Group), Interim Guidance Document on Cholera Surveillance (2017).

8. F. Qadri, A. I. Khan, A. S. G. Faruque, Y. A. Begum, F. Chowdhury, G. B. Nair, M. A. Salam, D. A. Sack, A.-M. Svennerholm, Enterotoxigenic *Escherichia coli* and *Vibrio cholerae* diarrhea, Bangladesh, 2004. *Emerg. Infect. Dis.* **11**, 1104–1107 (2005).

9. W. H. Chen, M. B. Cohen, B. D. Kirkpatrick, R. C. Brady, D. Galloway, M. Gurwith, R. H. Hall, R. A. Kessler, M. Lock, D. Haney, C. E. Lyon, M. F. Pasetti, J. K. Simon, F. Szabo, S. Tennant, M. M. Levine, Single-dose Live Oral Cholera Vaccine CVD 103-HgR Protects Against Human Experimental Infection With *Vibrio cholerae* O1 El Tor. *Clin. Infect. Dis.* **62**, 1329–1335 (2016).

10. A. S. Azman, S. A. Lauer, T. R. Bhuiyan, F. J. Luquero, D. T. Leung, S. T. Hegde, J. B. Harris, K. K. Paul, F. Khaton, J. Ferdous, J. Lessler, H.

Salje, F. Qadri, E. S. Gurley, Vibrio cholerae O1 transmission in Bangladesh: insights from a nationally representative serosurvey. *Lancet Microbe*. **1**, e336–e343 (2020).

11. R. R. German, L. M. Lee, J. M. Horan, R. L. Milstein, C. A. Pertowski, M. N. Waller, Guidelines Working Group Centers for Disease Control and Prevention (CDC), *MMWR Recomm. Rep.*, in press.

12. S. T. Hegde, E. C. Lee, A. Islam Khan, S. A. Lauer, M. T. Islam, T. Rahman Bhuiyan, J. Lessler, A. S. Azman, F. Qadri, E. S. Gurley, Clinical Cholera Surveillance Sensitivity in Bangladesh and Implications for Large-Scale Disease Control. *J. Infect. Dis.* **224**, S725–S731 (2021).

13. Cholera Roadmap Research Agenda, (available at <https://www.gtfcc.org/cholera-roadmap-research-agenda/>).

14. C. J. E. Metcalf, J. Farrar, F. T. Cutts, N. E. Basta, A. L. Graham, J. Lessler, N. M. Ferguson, D. S. Burke, B. T. Grenfell, Use of serological surveys to generate key insights into the changing global landscape of infectious disease. *Lancet*. **388**, 728–730 (2016).

15. A. M. Harris, M. S. Bhuiyan, F. Chowdhury, A. I. Khan, A. Hossain, E. A. Kendall, A. Rahman, R. C. LaRocque, J. Wrammert, E. T. Ryan, F. Qadri, S. B. Calderwood, J. B. Harris, Antigen-specific memory B-cell responses to Vibrio cholerae O1 infection in Bangladesh. *Infect. Immun.* **77**, 3850–3856 (2009).

16. M. L. Clements, M. M. Levine, C. R. Young, R. E. Black, Y. L. Lim, R. M. Robins-Browne, J. P. Craig, Magnitude, kinetics, and duration of vibriocidal antibody responses in North Americans after ingestion of *Vibrio cholerae*. *J. Infect. Dis.* **145**, 465–473 (1982).
17. M. S. Son, R. K. Taylor, Vibriocidal assays to determine the antibody titer of patient sera samples. *Curr. Protoc. Microbiol.* **Chapter 6**, Unit6A.3 (2011).
18. A. S. Azman, J. Lessler, F. J. Luquero, T. R. Bhuiyan, A. I. Khan, F. Chowdhury, A. Kabir, M. Gurwith, A. A. Weil, J. B. Harris, S. B. Calderwood, E. T. Ryan, F. Qadri, D. T. Leung, Estimating cholera incidence with cross-sectional serology. *Sci. Transl. Med.* **11** (2019), doi:10.1126/scitranslmed.aau6242.
19. B. F. Arnold, M. J. van der Laan, A. E. Hubbard, C. Steel, J. Kubofcik, K. L. Hamlin, D. M. Moss, T. B. Nutman, J. W. Priest, P. J. Lammie, Measuring changes in transmission of neglected tropical diseases, malaria, and enteric pathogens from quantitative antibody levels. *PLoS Negl. Trop. Dis.* **11**, e0005616 (2017).
20. J. Rosado, S. Pelleau, C. Cockram, S. H. Merklung, N. Nekkab, C. Demeret, A. Meola, S. Kerneis, B. Terrier, S. Fafi-Kremer, J. de Seze, T. Bruel, F. Dejardin, S. Petres, R. Longley, A. Fontanet, M. Backovic, I. Mueller, M. T. White, Multiplex assays for the identification of serological signatures of SARS-CoV-2 infection: an antibody-based diagnostic and machine learning study. *Lancet Microbe.* **2**, e60–e69 (2021).

21. G. P. Smits, P. G. van Gageldonk, L. M. Schouls, F. R. M. van der Klis, G. A. M. Berbers, Development of a Bead-Based Multiplex Immunoassay for Simultaneous Quantitative Detection of IgG Serum Antibodies against Measles, Mumps, Rubella, and Varicella-Zoster Virus. *Clinical and Vaccine Immunology*. **19** (2012), pp. 396–400.
22. P. J. Lammie, D. M. Moss, E. Brook Goodhew, K. Hamlin, A. Krolewiecki, S. K. West, J. W. Priest, Development of a new platform for neglected tropical disease surveillance. *Int. J. Parasitol.* **42**, 797–800 (2012).
23. A. Ayouba, A. Touré, C. Butel, A. K. Keita, F. Binetruy, M. S. Sow, V. Foulongne, E. Delaporte, M. Peeters, Development of a Sensitive and Specific Serological Assay Based on Luminex Technology for Detection of Antibodies to Zaire Ebola Virus. *J. Clin. Microbiol.* **55**, 165–176 (2017).
24. J. W. Priest, D. M. Moss, Measuring Cryptosporidium Serologic Responses by Multiplex Bead Assay. *Methods Mol. Biol.* **2052**, 61–85 (2020).
25. B. F. Arnold, H. M. Scobie, J. W. Priest, P. J. Lammie, Integrated Serologic Surveillance of Population Immunity and Disease Transmission. *Emerg. Infect. Dis.* **24**, 1188–1194 (2018).
26. S. M. Patel, M. A. Rahman, M. Mohasin, M. A. Riyadh, D. T. Leung, M. M. Alam, F. Chowdhury, A. I. Khan, A. A. Weil, A. Aktar, M. Nazim, R. C. LaRocque, E. T. Ryan, S. B. Calderwood, F. Qadri, J. B. Harris, Memory B cell responses to *Vibrio cholerae* O1 lipopolysaccharide are associated with

protection against infection from household contacts of patients with cholera in Bangladesh. *Clin. Vaccine Immunol.* **19**, 842–848 (2012).

27. A. Aktar, M. A. Rahman, S. Afrin, M. O. Faruk, T. Uddin, A. Akter, M. I. N. Sami, T. Yasmin, F. Chowdhury, A. I. Khan, D. T. Leung, R. C. LaRocque, R. C. Charles, T. R. Bhuiyan, A. Mandlik, M. Kelly, P. Kováč, P. Xu, S. B. Calderwood, J. B. Harris, F. Qadri, E. T. Ryan, O-Specific Polysaccharide-Specific Memory B Cell Responses in Young Children, Older Children, and Adults Infected with *Vibrio cholerae* O1 Ogawa in Bangladesh. *Clin. Vaccine Immunol.* **23**, 427–435 (2016).

28. A. S. Azman, K. E. Rudolph, D. A. T. Cummings, J. Lessler, The incubation period of cholera: a systematic review. *J. Infect.* **66**, 432–438 (2013).

29. E. T. Ryan, D. T. Leung, O. Jensen, A. A. Weil, T. R. Bhuiyan, A. I. Khan, F. Chowdhury, R. C. LaRocque, J. B. Harris, S. B. Calderwood, F. Qadri, R. C. Charles, Systemic, Mucosal, and Memory Immune Responses following Cholera. *Tropical Medicine and Infectious Disease.* **6**, 192 (2021).

30. R. C. Kauffman, T. R. Bhuiyan, R. Nakajima, L. M. Mayo-Smith, R. Rashu, M. R. Hoq, F. Chowdhury, A. I. Khan, A. Rahman, S. K. Bhaumik, L. Harris, J. T. O’Neal, J. F. Trost, N. H. Alam, A. Jasinskas, E. Dotsey, M. Kelly, R. C. Charles, P. Xu, P. Kováč, S. B. Calderwood, E. T. Ryan, P. L. Felgner, F. Qadri, J. Wrammert, J. B. Harris, Single-Cell Analysis of the Plasmablast Response to *Vibrio cholerae* Demonstrates Expansion of Cross-Reactive Memory B Cells. *MBio.* **7** (2016), doi:10.1128/mBio.02021-16.



31. R. C. Charles, R. Nakajima, L. Liang, A. Jasinskas, A. Berger, D. T. Leung, M. Kelly, P. Xu, P. Kováč, S. R. Giffen, Others, Plasma and mucosal immunoglobulin M, immunoglobulin A, and immunoglobulin G responses to the *Vibrio cholerae* O1 protein immunome in adults with cholera in Bangladesh. *J. Infect. Dis.* **216**, 125–134 (2017).
32. Manual, xMAP Antibody Coupling Kit User, English. *Luminex Corporation* (2014), (available at <https://www.luminexcorp.com/download/xmap-antibody-coupling-kit-user-manual/>).
33. L. L. van den Hoogen, J. Présumé, I. Romilus, G. Mondélus, T. Elismé, N. Sepúlveda, G. Stresman, T. Druetz, R. A. Ashton, V. Joseph, T. P. Eisele, K. E. S. Hamre, M. A. Chang, J. F. Lemoine, K. K. A. Tetteh, J. Boncy, A. Existe, C. Drakeley, E. Rogier, Quality control of multiplex antibody detection in samples from large-scale surveys: the example of malaria in Haiti. *Scientific Reports.* **10** (2020), , doi:10.1038/s41598-020-57876-0.
34. C. Ritz, F. Baty, J. C. Streibig, D. Gerhard, Dose-Response Analysis Using R. *PLoS One.* **10**, e0146021 (2015).
35. RStan, (available at <https://mc-stan.org/users/interfaces/rstan>).
36. H. Salje, D. A. T. Cummings, I. Rodriguez-Barraquer, L. C. Katzelnick, J. Lessler, C. Klungthong, B. Thaisomboonsuk, A. Nisalak, A. Weg, D. Ellison, L. Macareo, I.-K. Yoon, R. Jarman, S. Thomas, A. L. Rothman, T. Endy, S. Cauchemez, Reconstruction of antibody dynamics and infection histories to evaluate dengue risk. *Nature.* **557**, 719–723 (2018).

37. Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models [R package loo version 2.4.1] (2020) (available at <https://cran.r-project.org/web/packages/loo/index.html>).
38. M. N. Wright, S. Wager, P. Probst, Ranger: A fast implementation of random forests. *R package version 0.12.1* (2020) (available at <https://xscodex.com/imbs-hl/ranger>).
39. A. Altmann, L. Toloşi, O. Sander, T. Lengauer, Permutation importance: a corrected feature importance measure. *Bioinformatics*. **26**, 1340–1347 (2010).
40. C. Kennedy, University of California, Berkeley, Guide to SuperLearner (2017), (available at <https://cran.r-project.org/web/packages/SuperLearner/vignettes/Guide-to-SuperLearner.html>).
41. B. Goodrich, J. Gabry, I. Ali, S. Brilleman, rstanarm: Bayesian applied regression modeling via Stan. *R package version*. **2**, 1758 (2018).
42. Y. Chan, K. Fornace, L. Wu, B. F. Arnold, J. W. Priest, D. L. Martin, M. A. Chang, J. Cook, G. Stresman, C. Drakeley, Determining seropositivity-A review of approaches to define population seroprevalence when using multiplex bead assays to assess burden of tropical diseases. *PLoS Negl. Trop. Dis.* **15**, e0009457 (2021).
43. L. M. Mayo-Smith, J. K. Simon, W. H. Chen, D. Haney, M. Lock, C. E. Lyon, S. B. Calderwood, B. D. Kirkpatrick, M. Cohen, M. M. Levine, M. Gurwith, J. B. Harris, The Live Attenuated Cholera Vaccine CVD 103-HgR

Primes Responses to the Toxin-Coregulated Pilus Antigen TcpA in Subjects Challenged with Wild-Type *Vibrio cholerae*. *Clin. Vaccine Immunol.* **24** (2017), doi:10.1128/CVI.00470-16.

44. S. L. Karlsson, N. Thomson, A. Mutreja, T. Connor, D. Sur, M. Ali, J. Clemens, G. Dougan, J. Holmgren, M. Lebens, Retrospective Analysis of Serotype Switching of *Vibrio cholerae* O1 in a Cholera Endemic Region Shows It Is a Non-random Process. *PLoS Negl. Trop. Dis.* **10**, e0005044 (2016).

45. Getting Double the Data from Less Sample: An xMAP INTELLIFLEX® Dual Reporter Assay in Action. *Luminex Corporation* (2021), (available at <https://www.luminexcorp.com/blog/getting-double-the-data-from-less-sample-an-xmap-intelliflex-dual-reporter-assay-in-action/>).

## **CHAPTER 4: Conducting cholera serosurveillance in partially vaccinated populations**

Forrest K. Jones, Taufiq R. Bhuiyan, Damien Slater, Kian Hutt Vater, Ashraful I Khan, Ralph Tennier, Kennia Visieres, Louise Ivers, Regina C. LaRocque, Stephen B. Calderwood, Justin Lessler, Daniel T. Leung, Edward T. Ryan, Richelle C. Charles, Fahima Chowdhury, Firdausi Qadri, Jason B. Harris, Andrew S. Azman

### **Abstract**

Targeting mass oral cholera vaccination (OCV) campaigns to subnational areas with high incidence has been proposed to reduce global cholera burden and transmission. Serological surveillance could provide complementary information about *Vibrio cholerae* infections missed by clinical surveillance in these areas. Similar immune responses generated by vaccination and infection may lead to biased estimates of incidence from serological surveys after campaigns. Using serological data from cohorts of infected and vaccinated individuals, we characterized antibody dynamics in both groups and explored strategies to estimate seroincidence in partially vaccinated populations.

We tested 305 serum samples from 51 confirmed *V. cholerae* serogroup O1 cases and uninfected contacts enrolled in Bangladesh and 248 serum samples from 51 volunteers vaccinated with killed whole-cell cholera vaccine in Haiti for IgG, IgM, and IgA antibodies binding to 8 *V. cholerae* antigens using a

multiplex bead assay. Both vaccination and infection stimulated anti-Ogawa OSP and anti-Inaba OSP antibody responses, but only infection led to detectable rises in anti-CT-B and anti-TcpA antibodies. Classification models trained only to detect recent infections (using 3 IgG antibody measurements) often misclassified vaccinated individuals as recently infected except over short seroincidence periods (e.g. 45-day seroincidence). Through cross-validation, we found that classification models trained using 15 antibody measurements could partially differentiate samples from 1) adults infected in the last 120 days (sensitivity: 64%), 2) adults vaccinated but not recently infected (sensitivity: 81%), and 3) adults neither recently infected nor vaccinated (sensitivity: 79%). Using simulated cross-sectional serological surveys, we found that measuring additional serological markers or ascertaining vaccination status by other means (such as a questionnaire) could be used to accurately estimate seroincidence.

Estimating incidence over a shorter window of time, testing for additional antibodies, and ascertaining vaccination status are all feasible strategies for undertaking serological surveillance shortly after vaccination campaigns. Serological surveillance can be a viable epidemiologic tool in high cholera incidence areas regardless of vaccination coverage.

## **Introduction**

Cholera remains a global public health threat, with an estimated 95,000 deaths per year (1). The Global Task Force on Cholera Control's End Cholera 2030 Roadmap is based on highly focused disease prevention and control in subnational 'cholera hotspots' (2). Killed oral cholera vaccines (OCV) are one

recommended component of cholera prevention in hotspots as they can reduce severe cholera symptoms and limit transmission (3, 4). Clinical surveillance for cholera can be poor, especially in highly affected communities, with most infections being missed and many suspected cases being misattributed as cholera (5). Serological surveillance (i.e. serosurveillance) has been proposed to gain complementary information on cholera transmission and burden. However, as vaccination becomes more prevalent, it is unclear how to adapt serosurveillance methods to make meaningful inference on infection rates in partially-vaccinated populations.

Similarities between the immune response to *V. cholerae* infection and OCV have been compared in previous research (6). The two most used OCV in cholera endemic regions (Shanchol and Euvichol) contain five inactivated strains of *V. cholerae*, with no component of the cholera toxin (7). Both O1 (four strains) and O139 (one strain less commonly found in humans) serogroups are included; among the O1 strains, there are two Ogawa and two Inaba serotype strains. The vaccine is administered to individuals over the age of 1 year in two doses spaced at least two weeks apart. Mass campaigns generally last a short period of time, with each dose being administered to a population often over the course of a week or less (8). Vaccination has been shown to boost serological markers that have been proposed to identify previously infected individuals (9); seroincidence estimates from cross-sectional surveys conducted in partially vaccinated populations may be biased due to misclassification of vaccinated individuals as recently infected.

As has been shown for SARS-CoV-2 (10, 11) and dengue (12), measuring additional serological markers or differences in post-exposure antibody kinetics may be useful to distinguish vaccinated and infected individuals. Statistical adjustment of seroincidence estimates using information on vaccination status, either inferred through questionnaire data or serologic profiles, may help reduce biases in estimating seroincidence. Avoiding serological surveys shortly after vaccination campaigns, when vaccine-induced antibodies are at their highest, may reduce bias, but it remains unclear how long a wait is necessary. Additional data on the antibody kinetics following *V. cholerae* infection and/or vaccination are needed to evaluate these strategies.

In this study, we measured antibodies to 8 *V. cholerae* antigens in longitudinally collected serum samples after infection (up to 3 years) and vaccination (up to 1 year). We evaluated classification models trained on serological data to distinguish recently vaccinated, recently infected, or individuals who were neither recently vaccinated nor infected. Using statistical models trained on only three antibody markers to identify recent cholera cases, we estimated the risk of misclassifying recently vaccinated individuals as recently infected and how this varied with time since vaccination. Lastly, we simulated serological surveys to understand whether using additional serological data or identifying vaccinated individuals by other means (such as a questionnaire) could be used to adjust seroincidence estimates to account for misclassification of vaccinated individuals as recently infected.

## Methods

### Study Population

In this study we analyzed data from two cohorts: volunteers vaccinated with OCV (Shanchol) enrolled in Haiti (i.e. vaccinees) and confirmed cholera cases and their household contacts enrolled in Bangladesh.

In Bangladesh, as described previously, consenting patients (>1 year old) hospitalized at the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) Dhaka hospital with culture confirmed *V. cholerae* O1 were enrolled between 2006 and 2018 (13, 14). We utilized data from a previous analysis where we selected a sample of 51 confirmed cases (2 - 1083 days post symptom onset) and uninfected contacts enrolled in Bangladesh (305 serum samples in total).

In Haiti, serum samples were collected from healthy volunteers enrolled from the outpatient department at Saint Nicholas Hospital in St. Marc, Haiti, an urban center in the Artibonite Department. Individuals were excluded if previously given OCV, pregnant, or they had active gastrointestinal disorder within 7 days prior to enrollment. All individuals received two doses of OCV spaced 14 days apart. Children (i.e 1-17 years old) were enrolled during April 2013, while adults (i.e.  $\geq 18$  years) were enrolled in 2015 (May) and 2016 (January, March and April). While cholera cases were reported regularly in Haiti after being introduced in 2010, cases steadily decreased from 2011 through the time of these studies (15). Children had serum samples collected around days 0,



7, and 21 post first dose while adults had samples collected around 0, 7, 21, 44, 90, 180, 270, and 365 days post first dose.

We had access to samples from 98 child vaccinees followed up to 21 days post vaccination and 73 adult vaccinees followed up to 360 days. To conserve sera and laboratory reagents, we chose to select samples from 51 vaccinees. After limiting selection to individuals with more than 3 samples, we randomly selected five vaccinees  $\leq 5$  years, five vaccinees 6-9 years-old, five vaccinees 10-17 years old, and 36 vaccinees  $\geq 18$  years old. In total, there were 258 samples from 51 vaccinees. To limit the influence of boosted antibody responses from infection during the follow-up period, we removed any data points that were part of or after a greater than 2 fold-rise between measurements in vibriocidal (Ogawa serotype) titers  $>90$  days post initial infection. As all rises greater than 2-fold for vibriocidal Inaba titers followed a previous drop in titers below a detectable range (which were considered unlikely to be a part of reinfection pattern), only the Ogawa serotype was used for identifying reinfections.

#### Serological testing and data processing

Based on previous work on the immune response to *V. cholerae* infection (16–18), we selected a panel of eleven antigens to investigate with a multiplex bead assay (as described in Chapter 3). These included O1 serogroup Ogawa serotype O-specific polysaccharide (OSP, part of the LPS), O1 serogroup Inaba serotype OSP, CT-B, cholera toxin holotoxin (CT-H), Toxin co-regulated pilus subunit A (TcpA), *V. cholerae* cytolysin (VCC) (also known as hemolysin A), *V. cholerae* sialidase, and O139 serogroup OSP (*V. cholerae* O139 serogroup is

rarely detected as circulating, but is included in most vaccines). Additionally, heat labile enterotoxin subunit B (LT-B), and heat labile holo-enterotoxin (LT-H) (expressed during infection with enterotoxigenic *Escherichia coli* [ETEC] and have a high degree of homology with cholera toxin) were also selected. Influenza hemagglutinin 1 (Flu) was also included as a control antigen. All antigens were conjugated to Luminex magnetic beads.

Plate layouts were designed to include a dilution series (from pooled convalescent sera of culture confirmed *V. cholerae* O1 infection) and control wells, all of which were run in triplicate. Following the testing protocol, serum, beads, and secondary antibodies binding to IgG, IgA, and IgM were added to each well. Samples were run on a Luminex Flexmap 3D machine at Massachusetts General Hospital by one technician. Bead counts and median fluorescence intensity (MFI) values were exported from the Exponent software program. Plates were retested when more than half of the positive control dilutions had  $\geq 5$  antigens (excluding O139 OSP) with a coefficient of variation (calculated from triplicate MFI measurements) greater than 20%. For the analysis, any measurements with a bead count less than 30 were excluded ( $< 0.1\%$ ). MFI values were averaged across replicate wells. We calculated the Net MFI for each sample (i.e., MFI of sample - MFI of blank well, but censored at 10 FI units).

### Statistical Analyses

We fit hierarchical regression models for each marker to estimate the degree of antibody boosting post-infection and its decay rate after the boost for

each serological marker. We used a Bayesian framework with two components: a kinetic model and a measurement model. For the kinetic model, based on previous analyses of serological data (19), we assumed individuals had an instantaneous boost of antibodies 5 days post-infection or vaccination followed by exponential decay. We chose the time of boosting at 5 days post-infection or vaccination as most markers peaked for individuals peaked shortly afterwards, but all baseline samples of cases were less than 5 days post-infection. In the measurement model, we assumed random error was normally distributed (on the log-scale) and accounted for the fact that some observations were censored (e.g. titration data). We fit the models using Markov Chain Monte Carlo methods implemented in Stan (20, 21) with separate models for child vaccinees, adult vaccinees, child cases, and adult cases.

We investigated whether multiple serological markers could distinguish recently vaccinated (either 7, 21, or 44 days post first dose) from recently infected (<120 days) and those neither recently infected nor vaccinated. We considered baseline samples from cases and vaccinees, case samples collected >120 days after infection, and samples from a single uninfected household contact as representative of neither recent infection nor vaccination. We also limited our dataset for these analyses to include only adults given the short follow-up time of child vaccinees. First, we used multidimensional scaling (22, 23) to visualize the immunologic profiles of each group including all markers with a greater than a 4-fold boost from either vaccination or infection (as estimated by kinetic models). We then fit random forest models (24) using 15 antibody

measurements (anti-OSP, anti-CT-B, and anti-TcpA for all isotypes) and assessed accuracy using leave-one-individual-out cross validation (LOOCV, where each individual was in a single fold). Samples were classified based on which class received the highest number of votes. After estimating sensitivity and specificity of detecting vaccinated but uninfected individuals, we calculated the positive predictive value and negative predictive value at different levels of vaccine coverage (assuming a 120-day seroincidence of 5%).

We also identified periods where the risk of misclassifying vaccinated individuals as recent cases was the highest using previously recommended methods for estimating seroincidence (in Chapter 3). First, we trained random forest models on serological data (three IgG markers: anti-Ogawa OSP, anti-CT-B, and anti-Inaba OSP) from case and uninfected household contacts (using methods described in Chapter 3) and established a cut-off for a nominal 95% specificity in the unvaccinated population. We then predicted the serostatus (i.e. prediction of being recently infected based on serological measurements) of each sample of vaccinees to estimate the false positivity rate at different times since vaccination. For vaccinated adults, we also fit hierarchical logistic regression models to estimate the time-varying function of seropositivity for time since vaccination. Specifically, we assumed that the  $\text{logit}(\text{seropositivity})$  was a cubic polynomial function of (log-transformed) days since infection.

We simulated serological surveys that occurred after vaccination campaigns and compared estimated seroincidence with three methods. The three strategies we considered were 1) ignoring misclassification due to

vaccination, 2) using knowledge of vaccination status to adjust estimates, and 3) using additional serological markers to distinguish vaccinated individuals. Given our dataset of adult vaccinees had samples from around 7, 30, and 76 days post second dose, those were the times between vaccination campaign and survey we simulated using. We considered scenarios with different levels of vaccination coverage (25%, 50%, 75%) and 5 cases per 100 cumulative infection incidence during a 200-day infection window.

We simulated infection and vaccination (assuming that they were independent) status for 2000 individuals, with ten rounds of simulation each scenario. Using estimates of sensitivity and specificity from the LOOCV analysis, we simulated whether each adult was seropositive or seronegative based on their infection status, vaccination status, and the model that was used. We assumed that individuals who were both recently infected and recently vaccinated had the same serologic profile of those recently infected, but not recently vaccinated. We then implemented a Bayesian framework (Supplementary methods in Appendix C) using Markov Chain Monte Carlo methods ([20](#)) to jointly estimate seroincidence and vaccination coverage while incorporating uncertainty in estimates of sensitivity and specificity of classification models.

## Results

### Study population

We tested 248 samples from 51 vaccinees, all of whom received two doses of Shanchol spaced 14 days apart (Table 4.1). Most vaccinees (71%) were adults (i.e.  $\geq 18$  years old) while cases were on average younger with only 35% being adults. However, among children (i.e.  $< 18$  years old), vaccinees (20%) and cases (26%) had a similar proportion  $< 5$  years old. Overall, the proportion of female participants was similar among vaccines (33%) and cases (38%). Most cases (81%) had *V. cholerae* O1 serotype Ogawa isolated from their stool while the rest were the Inaba serotype.

All individuals in the vaccinated cohort had a baseline sample taken on the day of first dose vaccination. All child vaccinees had additional blood samples taken around days 7 and 21 after receiving their first dose. Samples were collected from most adult vaccinees between 6-8 days (100%), 20-21 days (97%), 41-43 days (61%), 87-95 days (94%), 173-188 days (42%), 221-224 days (22%) and 361-369 days (47%) (Figure C.1). Cases were followed up to 1083 days with blood samples collected periodically.

**Table 4.1. Individual characteristics of culture confirmed cholera patients and vaccinated individuals**

<b>Characteristics</b>	<b>Bangladeshi cholera cases* (n=48)</b>	<b>Haitian vaccinees (n=51)</b>
Age Group		
< 5 years (%)	8 (17)	3 (6)
5-9 years (%)	14 (29)	7 (14)
10-17 years (%)	9 (19)	5 (10)
18+ years (%)	17 (35)	36 (71)
Female (%)	18 (38)	17 (33)
<i>V. cholerae</i> O1 Ogawa isolated (%)	39 (81)	NA

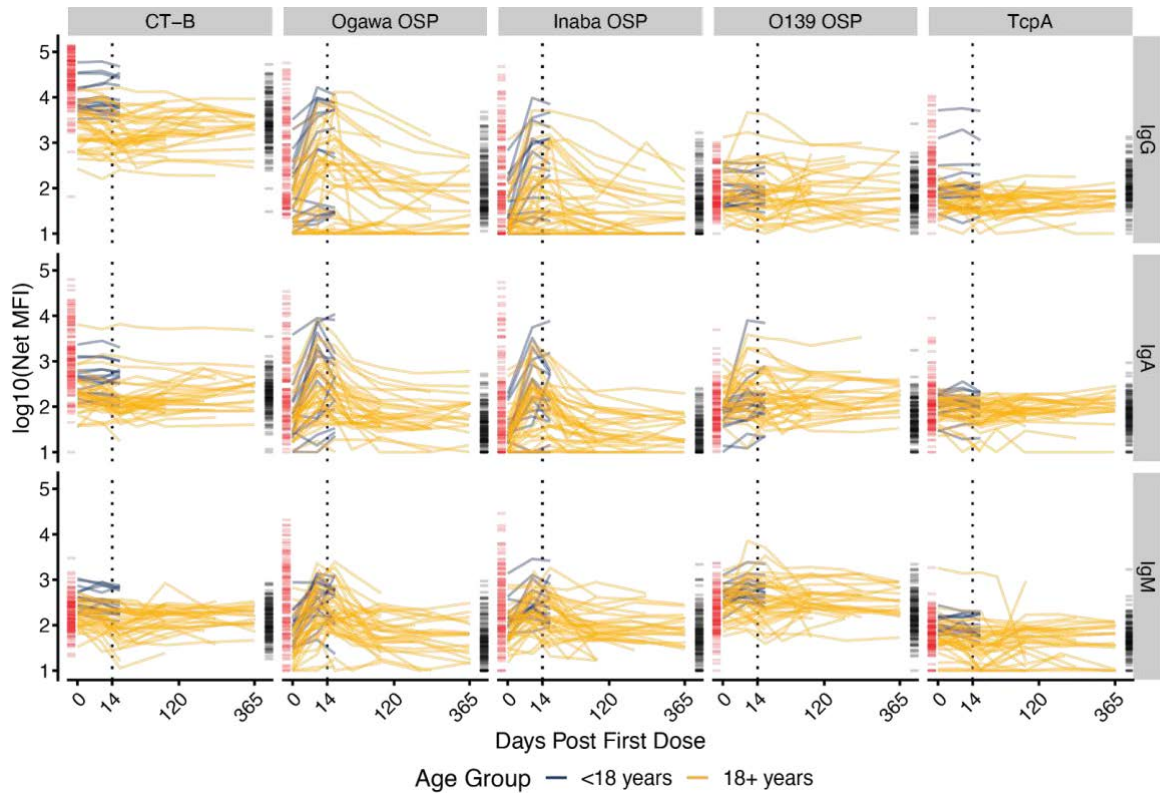
\* Serological data were also available for three uninfected household contacts of Bangladeshi cases enrolled with measurements taken at 2, 7, and 30 days after enrollment of the initial case

### Antibody kinetics of vaccinated volunteers and confirmed cholera cases

We first compared the earliest antibody measurements of vaccinees (i.e. day of first dose) and cases (i.e. between 2 and 4 days post-infection) for both adults and children. We observed little-to-no systematic differences between the baseline antibody measurements among adult cases and adult vaccinees. Child vaccinees, however, had elevated baseline levels of anti-CT-B and anti-OSP antibodies (ratio of the geometric mean Net MFI  $>2$ ) as compared to child cases across isotypes (Figure C.2). We considered this may be a sign of recent infection in some child vaccine recipients.

For vaccinees, anti-Ogawa OSP and anti-Inaba OSP antibodies (for all isotypes) rose steeply for most individuals after the first dose of vaccination (Figure 4.1, Figure C.3). Some antibody measurements were below the limit of detection (i.e.  $<1 \log_{10}(\text{Net MFI})$ ) for anti-Ogawa OSP and anti-Inaba IgG antibodies whereas nearly everyone had a detectable rise in IgA and IgM antibodies. The peak value in anti-Ogawa OSP and anti-Inaba OSP antibodies (across isotypes and age groups) for nearly every individual was around day 7 or day 21 (Figure C.4). Matching previous findings (25), we observed no apparent additional boost in antibodies from the second dose of vaccine provided two weeks after the first. As expected, we observed no rise in antibodies against CT-B after vaccination as CT-B is not included in Shanchol. Some vaccinees had a  $\geq 2$ -fold rise from baseline in anti-O139 OSP antibodies (IgG: 24%, IgA: 35%, IgM: 27%) at their peak.

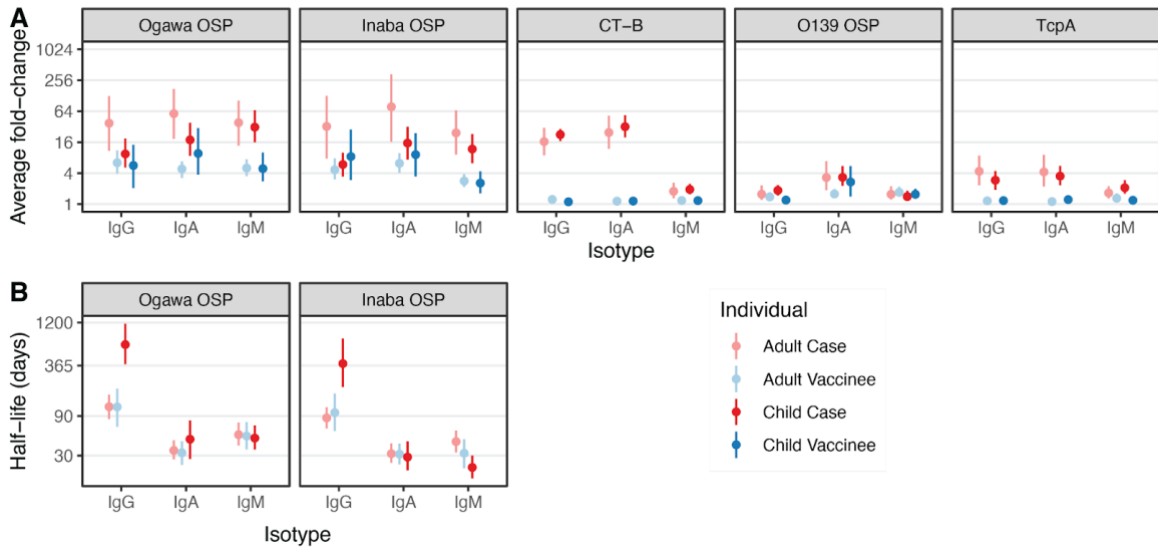




**Figure 4.1. Multiplex bead assay measurements of IgG, IgM, and IgA against CT-B, OSP, and TcpA antigens among Haitian vaccinated volunteers.**

Y-axis indicates the log (base 10) of the Net MFI. X-axis is square-root transformed. Each colored line indicates individual trajectories over time (dark blue: children <18 years, gold: adults  $\geq 18$  years). Rug plots show the antibody measurements from the cohort in Bangladesh (red: case measurements inside the 120-day infection window; black: uninfected household contacts and case measurements outside a 120-day infection window). The black dotted line indicates the timing of second dose vaccination.

To further describe the differences in the immune response, we fit kinetic models to different antibody measurements for both vaccinees and cases (Figure 4.2, C.5 & C.6, Table C.1). Among vaccinees, we estimated modest boosts (i.e., mean fold-change) for anti-Ogawa OSP antibodies for children (IgG: 5.6, IgA: 9.7, IgM: 4.9) and adults (IgG: 6.4, IgA: 4.8, IgM: 5.0). Among cases, we estimated much larger boosts for anti-Ogawa OSP antibodies for children (IgG: 9.5, IgA: 17.7, IgM: 31.3) and adults (IgG: 37.5, IgA: 57.5, IgM: 38.4). Higher boosts in anti-Inaba OSP antibodies were generally observed among cases than among vaccinees. Estimated average-fold change of anti-O139 OSP antibodies were all very low for both child vaccinees (IgG: 1.2, IgA: 2.7, IgM: 1.6) and adult vaccinees (IgG: 1.4, IgA: 1.6, IgM: 1.7). Among child and adult cases, anti-CT-B IgG (22.5 and 16.4 average fold-change) and anti-CT-B IgA (32 and 24.7 average fold-change) both rose, while effectively no boosting was observed among vaccinees. Small rises in anti-TcpA IgG and IgA (range of boosts: 2.9 - 4.4) were observed among cases, but not vaccinees. The median half-life for adult cases and adult vaccinees were very similar for anti-Ogawa OSP IgG (116 vs 116 days), IgA (35 vs 32 days), and IgM (54 vs 52 days) (as well as anti-Inaba OSP antibodies). Child cases had similar antibody decay rates as adult cases and vaccinees, though they had much slower decay of anti-Ogawa OSP IgG (654 days) and anti-Inaba OSP IgG (386 days). Among cases, the estimated half-life of anti-CT-B IgG (children: 123 days, adults: 146 days) and IgA (children: 27 days, adults: 33 days) was consistent across age groups (Table C.1).



**Figure 4.2. Estimated duration of half-life and average fold-change from exponential kinetic models for vaccinees and cases**

Points indicate the median parameter estimate for the average individual case (red: children, pink: adults) and vaccinee (dark blue: children, light blue: adults). Lines show the 95% credible interval (A). Estimated durations of half-life are not shown for anti-CT-B, anti-O139, and anti-TcpA due to little to no boosting among adult vaccinees and are not shown for child vaccinees due to short follow-up (B).

### *Distinguishing recently infected and vaccinated individuals using serological data*

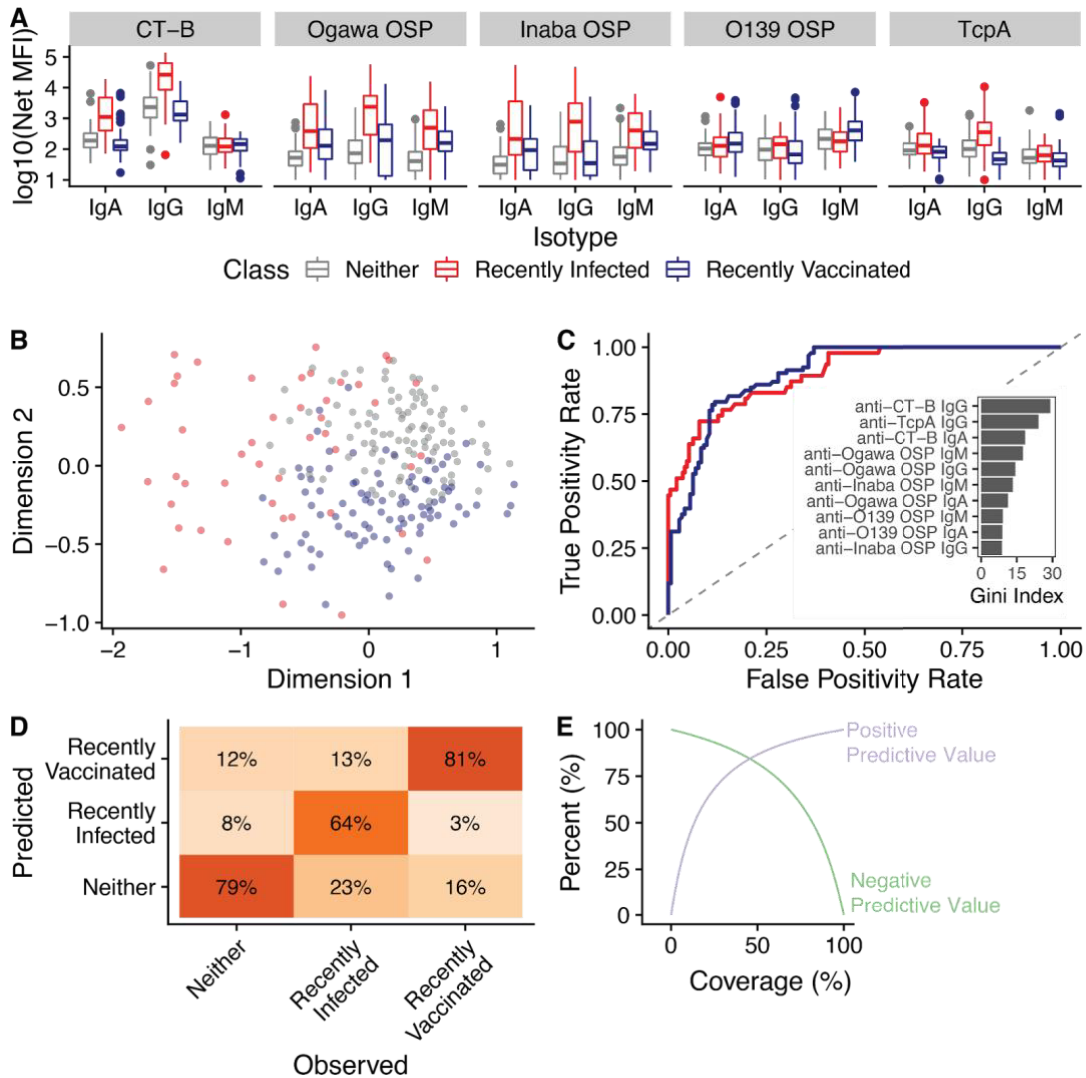
We investigated whether there were distinct immunological signatures that would allow us to distinguish between recently infected, recently vaccinated, and not infected or vaccinated individuals using 47 samples collected within 120 days of infection (17 adults), 93 samples collected within 44 days post first dose of vaccination (36 adults), and 94 samples in neither category (54 adults) (Figure 4.3A). The multidimensional scaling analysis demonstrated that these three populations were partially separated (though overlap remained), highlighting how combining serological data from several markers may be able to help differentiate these groups (Figure 4.3B).

Through cross-validation, we found that the random forest model trained on 15 serological markers could classify individuals into the categories of recently infected, recently vaccinated, or neither. Receiver-operator-characteristic curves demonstrated modest ability of the model to discriminate either recently infected or recently vaccinated samples from the other two classes (Figure 4.3C). The most influential markers were anti-CT-B IgG, anti-CT-B IgA, and anti-TcpA IgG, all of which were found to be stimulated by infection but not vaccination. Among markers stimulated by vaccination, anti-Ogawa OSP IgM was the most influential. Most samples from recently infected individuals (62%), recently vaccinated (80%) and neither infected nor recently vaccinated (79%) were correctly classified (Figure 4.3D). In a scenario with vaccination coverage of 50%, we estimated a specificity for identifying recently infected individuals of 94%. In a scenario with 5% infection incidence during the previous 120-days, we

estimated a specificity for identifying recently vaccinated individuals (but not recently infected) as 88%. Regardless of vaccine coverage, the positive and negative predictive values (of identifying recently vaccinated, but not infected) remained below 85% (Figure 4.3E).

#### *Misclassification of vaccinated individuals as recently infected*

We then considered what level of misclassification would occur using models trained on three antibody measurements (anti-Ogawa OSP, anti-Inaba OSP and anti-CT-B IgG) previously proposed for estimating seroincidence in unvaccinated populations (Chapter 3). After training random forest models using serological data from Bangladeshi cases and uninfected household contacts, we predicted the serostatus of vaccinated individuals to understand how often they would be misclassified. When utilizing a model trained to have 95% specificity, vaccinees were classified as recently infected at both baseline and at later time points ranging between 0% and 60% of the time (Table C.2). Among 15 baseline samples from child vaccinees a substantial proportion of samples were seropositive with models from each window (45-day: 7%, 120-day: 20%, 200-day: 27%, 300-day: 27%), potentially indicating some children were recently infected shortly before enrollment (or that the model has a greater tendency to misclassify children). As for adult vaccinees, baseline seropositivity for adult vaccinees was always below 5% (i.e. the expected false positivity rate) regardless of infection window. Among both child and adult vaccinees, the proportion seropositive increased at days 7 and 21 post first dose regardless of infection window (Table C.2).



**Figure 4.3. Serological measurements among recently infected, recently vaccinated, and neither recently vaccinated nor infected adults and classification model ability to distinguish between groups.**

Boxplots show the distributions for each antigen isotype combination (A). The two dimensions of the multidimensional scaling analysis are shown in the scatterplot (B). Receiver-operator characteristic curves describing the fifteen marker random forest classification model ability to identify recently infected (red) and recently vaccinated (blue) from the two other classes (C). Gini index importance metric for the top 10 markers is shown. A confusion matrix shows the proportion of samples predicted for each class during cross-validation (D). The line graph highlights the positive predictive and negative predictive value for identifying recently vaccinated (but not infected) individuals for different levels of vaccination coverage (E). The proportion of the population that has been recently infected is assumed to be 5% and that no individuals were both recently vaccinated and infected.

Among adult vaccinees, we modeled the seropositivity rate (i.e. the false positivity rate, FPR) as a function of time since the second dose (Figure 4.4A). When using the 45-day or 120-day infection window models, the FPR never increased above the expected FPR in unvaccinated populations (5%). We estimated that the maximum FPR occurred 3 days post second dose for the 200-day window (23.0% [95% CI: 16.5%-30.5%]) and 4 days post second dose (21.9% [95% CI: 15.9%-28.8%]) for the 300-day window, over four times the expected FPR in unvaccinated populations. When considering 200- and 300-day infection windows, the mean FPR among vaccinees remained above 5% for 152 and 157 days, respectively.

We also examined the accuracy of models trained using 15 markers to distinguish recently infected (<200 days), recently vaccinated individuals, and individuals in neither category, when limiting the time of vaccination to a single point in time, like in a vaccination campaign (Figure 4.4B). The model trained using samples 7 days post-second dose were reasonably sensitive and specific across categories, correctly classifying samples for most individuals in each class. Models trained using samples 21 or 76 day post-second dose performed similarly among recently infected and neither classes but were poor at classifying recently vaccinated individuals correctly (sensitivity < 35%).

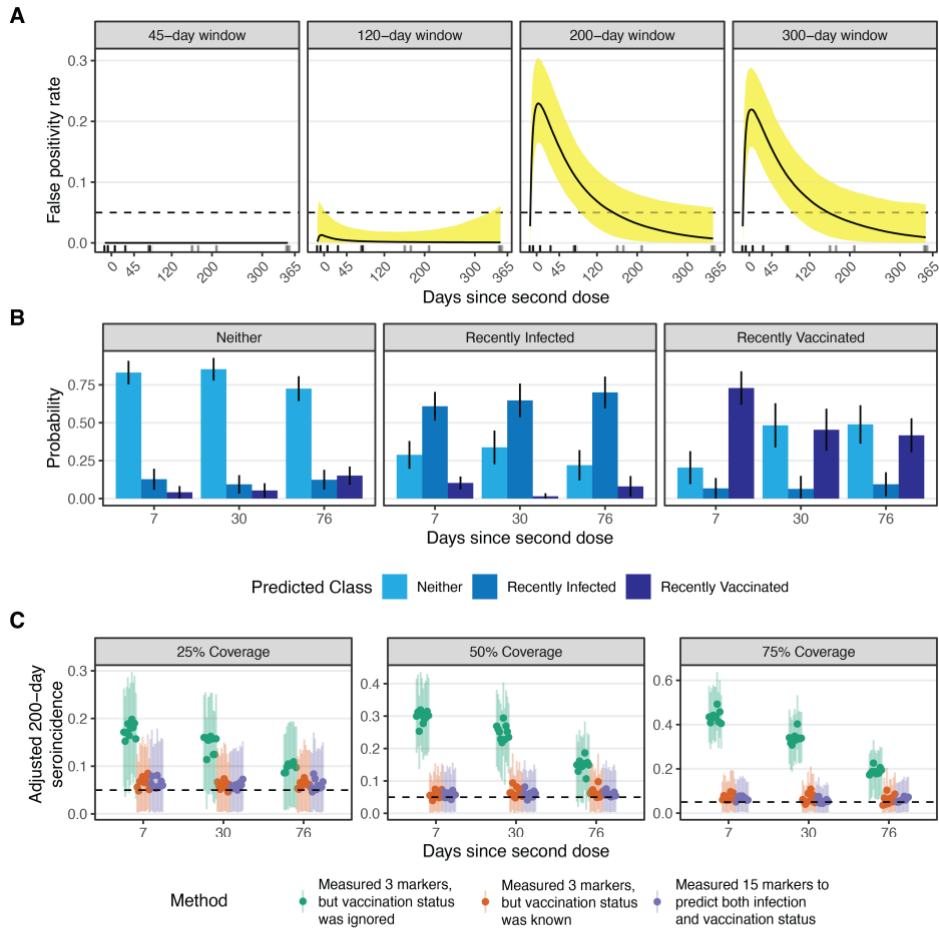
#### *Simulation of serological surveys and comparison of adjustment strategies*

Using simulated serological surveys conducted shortly after vaccination campaigns (7, 30, and 76 days after second dose), we observed how increased

misclassification led to greater bias in estimates of 200-day seroincidence. When vaccination status is ignored in the analysis and serological surveys occurred 7 days after the second vaccination campaign, we found that average seroincidence estimates were always much higher than the true incidence, regardless of coverage (Figure 4.4C). For example, at 25% vaccination coverage, the average seroincidence rate was estimated at 18 cases per 100, over >3.5 times the true rate of 5 cases per 100. This bias only increased when coverage was 50% (30 cases per 100) or 75% (44 cases per 100). We did find that bias decreased with additional time since the second dose, but estimated rates were still far above the true rate: at 90 days, the average seroincidence had dropped to 9.9 cases per 100 (25% coverage), 15 cases per 100 (50% coverage), and 19 cases per 100 (75% coverage).

We then examined how having either perfect information on individual vaccination status (e.g., perhaps as measured through a questionnaire) or testing for additional serological markers could reduce bias in adjusted estimates of 200-day seroincidence (Figure 4.4C, Table C.3). In the most extreme scenario (75% coverage and 7 days since the second dose), the average adjusted seroincidence was 6.4 cases per 100 when adjusting using known vaccination status and 6.0 cases per 100 when adjusting by testing for additional serological markers. We also found that strategies measuring serological markers generally led to unbiased estimates of vaccination coverage (Table C.4), though the uncertainty in those estimates increased dramatically at 30 and 76 days post second dose.





**Figure 4.4. Classification probability of samples by random forest models after vaccination and comparison of strategies for adjustment using simulated serological surveys.**

Each panel shows the false positivity rate (FPR) among vaccinated adults from models trained on three IgG serological markers (i.e. anti-Ogawa OSP, anti-Inaba OSP, anti-CT-B) with different infection windows (A). Solid lines indicate the median FPR on a given day and yellow shading indicates the 95% credible interval. The black dashed line indicates the nominal false positivity rate of 5% (i.e. specificity of 95%). Each panel shows the classification probabilities for different categories of random forest models trained to differentiate recently infected individuals (<200 days), recently vaccinated (for a given time since second dose), or neither (B). Black lines indicate the 95% confidence interval. Each panel shows a different scenario with varying levels of vaccination coverage (C). Points represent the median 200-day adjusted seroincidence estimate for 10 simulated surveys at each time point since the second dose of a vaccination campaign. Transparent lines represent the 95% credible interval. The black dashed line indicates the true incidence (5 cases per 100) during the 200-day infection window.

## Discussion

Despite similarities in the immune response between infection and vaccination, estimating cholera incidence through serological surveillance is feasible in partially vaccinated populations. We found that vaccination stimulates only a subset of the antibody repertoire of that generated by infection and to a lower degree. Classification models trained to identify recently infected individuals only misclassified vaccinated adults shortly after vaccination campaigns when using an infection window  $\geq 200$  days and only 3 IgG antibody measurements. We also found that both ascertaining vaccination status or measuring additional antibodies can be used to adjust for this misclassification.

Vaccination and infection have measurable differences in the antibody response generated that can be harnessed for serological surveillance. Though anti-Ogawa OSP and anti-Inaba OSP antibodies were generated among both cohorts, vaccinees generally had lower boosts overall. Anti-CT-B and anti-TcpA IgG and IgA antibodies were only stimulated by infection and not vaccination. We showed that classification models incorporating these markers were able to reasonably differentiate recently vaccinated and recently infected adults. However, this may be challenging to extend to children who may have higher levels of anti-CT-B antibodies (due to ETEC infection) and are less likely to generate anti-TcpA antibodies than adults (26). Lastly, we found that few anti-O139 OSP antibodies were generated from vaccination (matching previous findings (27)) and are unlikely to effectively differentiate vaccinees from infected individuals.

Stimulation of anti-OSP antibodies sometimes caused misclassification, but we found multiple ways to address this in the context of cross-sectional serological surveys. First, we found that models trained to identify individuals infected in the last 120 days rarely misclassify vaccinated individuals as recently infected. As the duration of most cholera outbreaks have been documented to be less than 120 days (28), estimating incidence in a cross-sectional survey shortly after an outbreak ends is unlikely to be impacted by vaccination campaigns. For situations where an infection window of 200 days is being considered, we found that misclassification can potentially be addressed either through measuring additional antibodies (e.g. additional isotypes and anti-TcpA antibodies) and/or ascertaining vaccination status (e.g. through a questionnaire). However, the adjustment we proposed assumes that all individuals are vaccinated within a short window of time and requires unbiased, precise estimates of sensitivity and specificity for the classification models. When starting a cross-sectional serosurvey, models to identify recent infections would ideally be validated with longitudinal serum samples from a population with similar cholera epidemiology.

Conducting cholera serosurveillance in vaccinated populations can provide meaningful inference and requires careful consideration about study goals. We showed how estimating seroincidence from cross-sectional studies can be managed for different infection windows in a partially vaccinated population. However, there may be challenges for estimating other parameters of interest such as the rate of detection by clinical surveillance. As vaccination can decrease the severity of cholera infection, there might differences in the detection

rate before and after the vaccination campaign. Also, we have shown the estimation of vaccination coverage using serological data is theoretically possible shortly after a vaccination campaign. This would require surveys occurring within a week or two after the campaign and measuring additional serological markers beyond anti-Ogawa OSP, anti-Inaba OSP, and anti-CT-B IgG (Table C.4). A questionnaire might be simpler for estimating vaccination coverage. Lastly, evaluation of interventions to prevent infection is theoretically possible using serological surveys but would likely require matching on important confounding factors.

The observed differences in serological data among cases and vaccinees in this study may be partially attributed to differences in the age distribution and history of cholera spread in the area shortly before sample collection. At baseline, we found that Haitian child vaccinees had higher levels of antibodies for *V. cholerae* antigens as compared to Bangladeshi child cases. These differences may be due to recent infection with *V. cholerae* as child vaccinees were enrolled in 2013, when rates of cholera were relatively high in Haiti. Given these differences and the relatively short follow-up time for child vaccinees, we limited our dataset to include only adults for much of our analysis. Though we hypothesize many of our findings comparing adult vaccinees and cases are broadly applicable, future research where both cases and vaccinees of all ages come from the same target population is needed to confirm them.

Our study population only contained individuals who were infected with *V. cholerae* or received two doses of OCV; there are several other combinations of

exposures that could exist in a target population and would impact serological measurements. ETEC infection is prevalent in many locations where cholera is, and stimulates antibodies that bind to both LT and CT. Further research is needed to understand how many ETEC infections may be misclassified as *V. cholerae* infection from models trained on serological data. Given the frequent co-occurrence of the two pathogens in the same settings (29), serosurveillance systems ought to be designed to estimate seroincidence for both together using anti-CT antibodies as well as pathogen-specific antibodies. Additionally, all vaccinees in our study received two doses of oral cholera vaccine exactly two weeks apart. In the context of real-world campaigns, some vaccinees will only receive one dose (potentially during the first or second round) or will receive two doses with a different time between doses. As more and more individuals are vaccinated, future serological studies will allow us to capture post-infection kinetics of individuals with a history of vaccination. Challenge studies, where exposure can be directly controlled, might also be well-suited for gathering serological data on individuals with different combinations of exposures.

Cholera serological surveillance provides valuable, complementary information in partially vaccinated populations. Expansion of cholera serosurveillance to new settings should include partially vaccinated populations

## **Acknowledgments**

We thank Rachel Mills for her help with laboratory analyses. We thank the Infectious Disease Dynamics Group at Johns Hopkins University for insight on the statistical analysis. We are grateful to Dr. Slavomír Bystrický Institute of

Chemistry, Slovak Academy of Sciences, Bratislava, Slovak Republic for the provision of *V. cholerae* O139 OSP. We acknowledge the study participants and families who consented to enroll into this study. *Funding*: This research was supported through programs funded by the National Institutes of Health, including the National Institute of Allergy and Infectious Diseases including R01 AI137164 (JBH, RCC), R01 AI106878 (ETR, FQ), U01 AI058935, U01 HD39165 (SBC, FQ, ETR), R01 AI135115 (DTL, ASA), the Fogarty International Center, Training Grant in Vaccine Development and Public Health (TW005572 [RB, MK]), and Emerging Global Fellowship Award TW010362 (TRB), and the Intramural Research Program of the NIH and NIDDK (PX and PK). We are grateful to the Governments of Bangladesh, Canada, Sweden and the UK for providing core/unrestricted support to icddr,b.

## References

1. Roadmap 2030, (available at <https://www.gtfcc.org/about-gtfcc/roadmap-2030/>).
2. E. C. Lee, A. S. Azman, J. Kaminsky, S. M. Moore, H. S. McKay, J. Lessler, The projected impact of geographic targeting of oral cholera vaccination in sub-Saharan Africa: A modeling study. *PLoS Med.* **16**, e1003003 (2019).
3. Q. Bi, E. Ferreras, L. Pezzoli, D. Legros, L. C. Ivers, K. Date, F. Qadri, L. Digilio, D. A. Sack, M. Ali, J. Lessler, F. J. Luquero, A. S. Azman, Oral Cholera Vaccine Working Group of The Global Task Force on Cholera Control,

Protection against cholera from killed whole-cell oral cholera vaccines: a systematic review and meta-analysis. *Lancet Infect. Dis.* **17**, 1080–1088 (2017).

4. A. S. Azman, J. Rumunu, A. Abubakar, H. West, I. Ciglenecki, T. Helderman, J. F. Wamala, O. de la R. Vázquez, W. Perea, D. A. Sack, D. Legros, S. Martin, J. Lessler, F. J. Luquero, Population-Level Effect of Cholera Vaccine on Displaced Populations, South Sudan, 2014. *Emerg. Infect. Dis.* **22**, 1067–1070 (2016).

5. S. T. Hegde, E. C. Lee, A. Islam Khan, S. A. Lauer, M. T. Islam, T. Rahman Bhuiyan, J. Lessler, A. S. Azman, F. Qadri, E. S. Gurley, Clinical Cholera Surveillance Sensitivity in Bangladesh and Implications for Large-Scale Disease Control. *J. Infect. Dis.* **224**, S725–S731 (2021).

6. D. T. Leung, M. A. Rahman, M. Mohasin, S. M. Patel, A. Aktar, F. Khanam, T. Uddin, M. A. Riyadh, A. Saha, M. M. Alam, F. Chowdhury, A. I. Khan, R. Charles, R. LaRocque, J. B. Harris, S. B. Calderwood, F. Qadri, E. T. Ryan, Memory B cell and other immune responses in children receiving two doses of an oral killed cholera vaccine compared to responses following natural cholera infection in Bangladesh. *Clin. Vaccine Immunol.* **19**, 690–698 (2012).

7. S. Kabir, Critical analysis of compositions and protective efficacies of oral killed cholera vaccines. *Clin. Vaccine Immunol.* **21**, 1195–1205 (2014).

8. L. Pezzoli, Global oral cholera vaccine use, 2013–2018. *Vaccine.* **38**, A132–A140 (2020).

9. A. S. Azman, J. Lessler, F. J. Luquero, T. R. Bhuiyan, A. I. Khan, F. Chowdhury, A. Kabir, M. Gurwith, A. A. Weil, J. B. Harris, S. B. Calderwood, E. T. Ryan, F. Qadri, D. T. Leung, Estimating cholera incidence with cross-sectional serology. *Sci. Transl. Med.* **11** (2019), doi:10.1126/scitranslmed.aau6242.
10. R. Assis, A. Jain, R. Nakajima, A. Jasinskas, S. Khan, A. Palma, D. M. Parker, A. Chau, S. Hosseinian, M. Vasudev, C. Au, K. Powers, P. S. Birring, B. Chin, R. Andary, J. M. Obiero, D. Tifrea, A. Leung, C. Grabar, F. Muqolli, G. Khalil, J. C. Escobar, J. Ventura, D. Huw Davies, B. Albala, B. Boden-Albala, S. Schubl, P. L. Felgner, Specimen Collection Group, Distinct SARS-CoV-2 antibody reactivity patterns elicited by natural infection and mRNA vaccination. *npj Vaccines.* **6** (2021), , doi:10.1038/s41541-021-00396-3.
11. A. Dörschug, H. Frickmann, J. Schwanbeck, E. Yilmaz, K. Mese, A. Hahn, U. Groß, A. E. Zautner, Comparative Assessment of Sera from Individuals after S-Gene RNA-Based SARS-CoV-2 Vaccination with Spike-Protein-Based and Nucleocapsid-Based Serological Assays. *Diagnostics.* **11** (2021), p. 426.
12. E. J. M. Nascimento, J. K. George, M. Velasco, M. I. Bonaparte, L. Zheng, C. A. DiazGranados, E. T. A. Marques, J. W. Huleatt, Development of an anti-dengue NS1 IgG ELISA to evaluate exposure to dengue virus. *J. Virol. Methods.* **257**, 48–57 (2018).
13. S. M. Patel, M. A. Rahman, M. Mohasin, M. A. Riyadh, D. T. Leung, M. M. Alam, F. Chowdhury, A. I. Khan, A. A. Weil, A. Aktar, M. Nazim, R. C. LaRocque, E. T. Ryan, S. B. Calderwood, F. Qadri, J. B. Harris, Memory B cell responses to *Vibrio cholerae* O1 lipopolysaccharide are associated with



protection against infection from household contacts of patients with cholera in Bangladesh. *Clin. Vaccine Immunol.* **19**, 842–848 (2012).

14. A. Aktar, M. A. Rahman, S. Afrin, M. O. Faruk, T. Uddin, A. Akter, M. I. N. Sami, T. Yasmin, F. Chowdhury, A. I. Khan, D. T. Leung, R. C. LaRocque, R. C. Charles, T. R. Bhuiyan, A. Mandlik, M. Kelly, P. Kováč, P. Xu, S. B. Calderwood, J. B. Harris, F. Qadri, E. T. Ryan, O-Specific Polysaccharide-Specific Memory B Cell Responses in Young Children, Older Children, and Adults Infected with *Vibrio cholerae* O1 Ogawa in Bangladesh. *Clin. Vaccine Immunol.* **23**, 427–435 (2016).

15. E. C. Lee, D. L. Chao, J. C. Lemaitre, L. Matrajt, D. Pasetto, J. Perez-Saez, F. Finger, A. Rinaldo, J. D. Sugimoto, M. Elizabeth Halloran, I. M. Longini, R. Ternier, K. Vissieres, A. S. Azman, J. Lessler, L. C. Ivers, Achieving coordinated national immunity and cholera elimination in Haiti through vaccination: a modelling study. *The Lancet Global Health.* **8** (2020), pp. e1081–e1089.

16. E. T. Ryan, D. T. Leung, O. Jensen, A. A. Weil, T. R. Bhuiyan, A. I. Khan, F. Chowdhury, R. C. LaRocque, J. B. Harris, S. B. Calderwood, F. Qadri, R. C. Charles, Systemic, Mucosal, and Memory Immune Responses following Cholera. *Tropical Medicine and Infectious Disease.* **6**, 192 (2021).

17. R. C. Kauffman, T. R. Bhuiyan, R. Nakajima, L. M. Mayo-Smith, R. Rashu, M. R. Hoq, F. Chowdhury, A. I. Khan, A. Rahman, S. K. Bhaumik, L. Harris, J. T. O’Neal, J. F. Trost, N. H. Alam, A. Jasinskas, E. Dotsey, M. Kelly, R. C. Charles, P. Xu, P. Kováč, S. B. Calderwood, E. T. Ryan, P. L. Felgner, F.

Qadri, J. Wrammert, J. B. Harris, Single-Cell Analysis of the Plasmablast Response to *Vibrio cholerae* Demonstrates Expansion of Cross-Reactive Memory B Cells. *MBio*. **7** (2016), doi:10.1128/mBio.02021-16.

18. R. C. Charles, R. Nakajima, L. Liang, A. Jasinskas, A. Berger, D. T. Leung, M. Kelly, P. Xu, P. Kováč, S. R. Giffen, Others, Plasma and mucosal immunoglobulin M, immunoglobulin A, and immunoglobulin G responses to the *Vibrio cholerae* O1 protein immunome in adults with cholera in Bangladesh. *J. Infect. Dis.* **216**, 125–134 (2017).

19. H. Salje, D. A. T. Cummings, I. Rodriguez-Barraquer, L. C. Katzelnick, J. Lessler, C. Klungthong, B. Thaisomboonsuk, A. Nisalak, A. Weg, D. Ellison, L. Macareo, I.-K. Yoon, R. Jarman, S. Thomas, A. L. Rothman, T. Endy, S. Cauchemez, Reconstruction of antibody dynamics and infection histories to evaluate dengue risk. *Nature*. **557**, 719–723 (2018).

20. RStan, (available at <https://mc-stan.org/users/interfaces/rstan>).

21. Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models [R package loo version 2.4.1] (2020) (available at <https://cran.r-project.org/web/packages/loo/index.html>).

22. J. de Leeuw, P. Mair, Multidimensional Scaling Using Majorization: SMACOF in R. *J. Stat. Softw.* **31**, 1–30 (2009).

23. S. Ashkiani, Dimensionality reduction and data visualization using MDS- SMACOF package in R, (available at <https://rstudio-pubs->

static.s3.amazonaws.com/246348\_b31bca1e4be04bb395825dc6a00de364.html)

24. M. N. Wright, S. Wager, P. Probst, Ranger: A fast implementation of random forests. *R package version 0. 12. 1* (2020) (available at <https://xscodex.com/imbs-hl/ranger>).

25. F. Chowdhury, A. Akter, T. R. Bhuiyan, I. Tauheed, S. Teshome, A. Sil, J. Y. Park, Y. Chon, J. Ferdous, S. R. Basher, F. Ahmed, M. Karim, M. M. Ahasan, M. R. Mia, M. M. I. Masud, A. W. Khan, M. Billah, Z. Nahar, I. Khan, A. G. Ross, D. R. Kim, M. M. R. Ashik, L. Digilio, J. Lynch, J.-L. Excler, J. D. Clemens, F. Qadri, A non-inferiority trial comparing two killed, whole cell, oral cholera vaccines (Cholvax vs. Shanchol) in Dhaka, Bangladesh. *Vaccine*. **40**, 640–649 (2022).

26. L. M. Mayo-Smith, J. K. Simon, W. H. Chen, D. Haney, M. Lock, C. E. Lyon, S. B. Calderwood, B. D. Kirkpatrick, M. Cohen, M. M. Levine, M. Gurwith, J. B. Harris, The Live Attenuated Cholera Vaccine CVD 103-HgR Primes Responses to the Toxin-Coregulated Pilus Antigen TcpA in Subjects Challenged with Wild-Type *Vibrio cholerae*. *Clin. Vaccine Immunol.* **24** (2017), doi:10.1128/CVI.00470-16.

27. V. Raghava Mohan, S. Raj, M. S. Dhingra, N. Aloysia D’Cor, A. P. Singh, T. Saluja, D. R. Kim, V. J. Midde, Y. Kim, S. Vemula, S. K. Narla, B. Sah, M. Ali, Safety and immunogenicity of a killed bivalent (O1 and O139) whole-cell oral cholera vaccine in adults and children in Vellore, South India. *PLoS One*. **14**, e0218033 (2019).

28. Q. Zheng, F. J. Luquero, I. Ciglenecki, J. F. Wamala, A. Abubakar, P. Welo, M. Hussen, M. Wossen, S. Yennan, A. Keita, J. Lessler, A. S. Azman, E. C. Lee, Cholera outbreaks in sub-Saharan Africa during 2010-2019: A Descriptive Analysis. *bioRxiv* (2021), , doi:10.1101/2021.10.25.21265347.

29. F. Qadri, A. I. Khan, A. S. G. Faruque, Y. A. Begum, F. Chowdhury, G. B. Nair, M. A. Salam, D. A. Sack, A.-M. Svennerholm, Enterotoxigenic *Escherichia coli* and *Vibrio cholerae* diarrhea, Bangladesh, 2004. *Emerg. Infect. Dis.* 11, 1104–1107 (2005).

## CHAPTER 5: Conclusion

Serological data, when measured in the context of surveillance or epidemiologic studies, can provide insight on the spread of, burden of, and immunity against infectious diseases at the population-level. However, before being able to make that inference, investigators need a clear understanding of what their assays are measuring. In this dissertation, I utilized datasets of longitudinally collected serum samples from laboratory confirmed cases to understand the dynamics of antibodies over time since infection and assess the accuracy of assays to detect previous infection with either SARS-CoV-2 or *V. cholerae*.

In Chapter 2, I showed how anti-RBD IgG antibodies detected through an in-house ELISA could provide valuable insight for detecting previous infection, with high levels of sensitivity (95%, after two weeks post symptom onset) and specificity (100%). The most important implication of this finding is that seroprevalence studies could provide valuable understanding of the incidence of SARS-CoV-2 infection as it was spreading across the world and could even be undertaken in areas with low-incidence given our high specificity. Such studies have been used to estimate important parameters (such as the infection fatality ratio and asymptomatic proportion) and to calibrate mathematical models (1–3). Additionally, I showed how SARS-CoV-2 infection did not lead to robust rises in antibodies of other human coronaviruses, indicating that issues concerning cross-reactivity (for this antigen) could likely be ignored.

Findings from Chapter 2 also have already made direct impacts for several scientific studies. For example, this work showed how anti-RBD IgG as measured from dried blood spots could be leveraged for serological surveys. Our team was able to use these data to estimate the number of infections missed by clinical surveillance during a serological survey in South Sudan (4). Additionally, our parameter estimates of the time required for seroconversion parameters can be used for statistical models attempting to incorporate dynamics in their estimates of incidence. For example, one study which attempted to estimate the cumulative incidence of SARS-CoV-2 infection in New York City and Connecticut used our estimate of the time to seroconversion in their model framework (5).

Finally, Chapter 2 also contained valuable insights on the dynamics of IgM and IgA, showing how both seroconversion and seroreversion occurred for both isotypes against RBD. This may mean that development of a seroincidence assay is possible for SARS-CoV-2, which may become more important as more and more people already have generated IgG antibodies to the spike protein, either through infection or vaccination. However, a new dataset would have to be generated among a non-immunologically naive population to develop an assay to detect recent infections. Additionally, other antigens (such as spike proteins from different variants or the nucleocapsid) might need to be considered to develop such an assay.

In Chapter 3, I demonstrated that the multiplex bead assay had similar performance at identifying recently infected individuals as traditional methods (that require the resource intensive vibriocidal assay). Multiplex bead assays are

already expanding to areas with neglected diseases like cholera (6). Adding *V. cholerae* antigens to existing multiplex bead assay panels may be far simpler than implementing vibriocidal methods. Additionally, smaller amounts of serum would be required for the multiplex bead assay, potentially meaning that valuable serum samples can be tested for more pathogens besides *V. cholerae*. Thus, these findings could help expand the number of areas worldwide with cholera serosurveillance. Before serosurveillance is implemented in an area (regardless of the assay used), careful consideration is needed to ensure the assay's false-positivity rate is lower than incidence in the area. Otherwise, any estimate of seroincidence provided may entirely be due to assay misclassification rather than true incident infections. Thus, areas with explosive outbreaks or high levels of endemic transmission (ideally with strong clinical surveillance and laboratory confirmation) should be prioritized first.

Expansion of serosurveillance systems for cholera has the potential to sharpen our understanding of the magnitude of spread and burden of the disease. Given the dependence on counting suspected cholera cases and infrequent laboratory confirmation, it remains largely unknown how common infection with *V. cholerae* is in the areas where it spreads. Serosurveillance may help identify areas with poor case-based surveillance that may not historically report cholera, but in fact have high levels of transmission. Serosurveillance could be used as a tool to monitor or evaluate programs (such as those geared towards improved water systems, sanitation or hygiene) that attempt to halt the transmission of cholera. Such evaluations might be accomplished through panel

surveys that estimate changes in incidence before and after an intervention has taken place or by comparing areas that did and did not receive an intervention. Such strategic information could aid efforts to reduce cholera morbidity and mortality as proposed by the Global Task Force on Cholera Control (7).

In Chapter 4, I found that cross-sectional serological surveys to estimate incidence could still be implemented in areas with cholera vaccination campaigns despite cross-reactive immune responses from vaccination. Given the expanding number of areas of the world with oral cholera vaccination campaigns, this is essential for serosurveillance to remain a viable tool. However, I did find that additional care needs to be taken when wanting to estimate incidence over the course of infection windows longer than 120 days. When attempting to estimate incidence over the course of such windows, the safest way for investigators to avoid misclassification of vaccinated individuals is to wait to conduct the survey such that the infection window does not contain the time of when the campaign has occurred. This may be convenient if the goal is to estimate incidence during the post-vaccination period.

However, waiting for antibodies (generated by vaccination) to wane may interfere with other study goals such as attempting to estimate incidence during an outbreak where a reactive campaign has occurred. In this situation, I found that measuring additional types of antibodies (such as increasing the number of antigens or isotypes measured) or ascertaining vaccination status independently (e.g., through a questionnaire) provides sufficient information to estimate incidence. Conducting serosurveillance in partially vaccinated populations still



requires additional research on individuals who have been both vaccinated and infected; understanding the antibody dynamics of these individuals will be key. Lastly, I found that few anti-O139 OSP antibodies were generated after vaccination against cholera, fitting with findings from past studies (8). As a result, anti-O139 OSP antibodies were not useful for distinguishing recently vaccinated from recently infected individuals and may not be worth measuring for future serosurveillance efforts.

Demonstrating antibody measurements to be correlates (or surrogates (9)) of protection for SARS-CoV-2 or *V. cholerae* was not a focus of this dissertation. There are several methodological challenges to establish them (even for diseases as well-studied as measles (10)); they are often investigated in the context of vaccine trials, observational studies, or challenge studies. In prior studies, serum antibodies that neutralize SARS-CoV-2 have been shown to correlate with protection against COVID-19 (11, 12). As new variants have evolved, especially those with immune escape properties, titers measured with neutralization assays using prior strains may have less use. Design of serosurveillance systems for SARS-CoV-2 (or any quickly evolving pathogen) should consider what variants are of greatest interest to inform decision of which strains of the virus should be used in neutralization assays for serosurveillance. As for *V. cholerae*, serum antibodies as measured by the vibriocidal assay have been shown to be imperfect correlates of protection against infection, with many individuals with high titers still being infected (13). Since *V. cholerae* infection occurs in the intestine, this assay does not directly measure antibodies involved

in mechanisms of immunity. Further investigation of antibodies measured from saliva or stool may hold promise as they might correlate well with those involved with mucosal immunity in the intestine.

Even if these correlates are only partially associated with protection, they might still be worth measuring in the context of serosurveillance activities. If the correlate is a marker of protection from infection, measurement at the population-level could provide useful insight as to whether widespread susceptibility is driving transmission as opposed to changes in behavior, climate factors, or interventions. Additionally, these data could help parameterize forecasting models of infectious disease, predict outbreaks before they begin, demonstrate how pathogens evolve in response to population-immunity, and drive important policy decisions (such as closing schools or instituting mask mandates). If the correlate is a marker of protection from disease, serosurveillance could identify which populations have the most vulnerable individuals. We can then target resources to those groups to help prevent illness in the first place (e.g. booster campaigns) or ensure treatment would be quickly available if needed.

In this dissertation, I applied laboratory, epidemiological, and statistical methods to answer questions about feasibility of serosurveillance for two biologically unrelated pathogens. Serosurveillance provides an opportunity to develop systems for monitoring several pathogens simultaneously rather than having separate systems each focusing on choice pathogens (14). Often some of the biggest challenges to implementing seroepidemiologic studies occur in the acquisition of participant specimens: issues can arise with acceptance,

enrollment, sample collection, transport, and storage. Proposals to start a “World Serology Bank” may be the path forward for addressing these challenges (15). With access to banked samples, multiplex bead assays could be leveraged to characterize antibody levels at the population-level for many pathogens while using less serum and being less costly than traditional assays (16).

With improvements to our understanding of incidence of infection and immunity to future threats, decisions affecting public health have the potential to be more informed than ever before. As generating serological data has become easier, more consideration of how serosurveillance can be combined with other forms of surveillance is needed. Serosurveillance is not a panacea to the challenges of imperfect information for monitoring infectious diseases. Rather, it is an important piece of the puzzle that has been underutilized for too many diseases for too long.

## References

1. S. Chen, J. A. Flegg, L. J. White, R. Aguas, Levels of SARS-CoV-2 population exposure are considerably higher than suggested by seroprevalence surveys, , doi:10.1101/2021.01.08.21249432.

2. R. Subramanian, Q. He, M. Pascual, Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2019716118.
3. G. Meyerowitz-Katz, L. Merone, A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. *International Journal of Infectious Diseases.* **101** (2020), pp. 138–148.
4. K. E. Wiens, P. N. Mawien, J. Rumunu, D. Slater, F. K. Jones, S. Moheed, A. Caflisch, B. K. Bior, I. A. Jacob, R. L. Lako, A. G. Guyo, O. O. Olu, S. Maleghemi, A. Baguma, J. J. Hassen, S. K. Baya, L. Deng, J. Lessler, M. N. Demby, V. Sanchez, R. Mills, C. Fraser, R. C. Charles, J. B. Harris, A. S. Azman, J. F. Wamala, Seroprevalence of severe acute respiratory syndrome Coronavirus 2 IgG in Juba, South Sudan, 2020. *Emerg. Infect. Dis.* **27**, 1598–1606 (2021).
5. K. Shioda, M. S. Y. Lau, A. N. M. Kraay, K. N. Nelson, A. J. Siegler, P. S. Sullivan, M. H. Collins, J. S. Weitz, B. A. Lopman, Estimating the Cumulative Incidence of SARS-CoV-2 Infection and the Infection Fatality Ratio in Light of Waning Antibodies. *Epidemiology.* **32**, 518–524 (2021).
6. Y. Chan, K. Fornace, L. Wu, B. F. Arnold, J. W. Priest, D. L. Martin, M. A. Chang, J. Cook, G. Stresman, C. Drakeley, Determining seropositivity-A review of approaches to define population seroprevalence when using multiplex bead assays to assess burden of tropical diseases. *PLoS Negl. Trop. Dis.* **15**, e0009457 (2021).

7. World Health Organization. Global Task Force on Cholera Control, *Ending Cholera, a Global Roadmap to 2030* (World Health Organization, 2017).
8. Y. O. Baik, S. K. Choi, R. M. Olveda, R. A. Espos, A. D. Ligsay, M. B. Montellano, J. S. Yeam, J. S. Yang, J. Y. Park, D. R. Kim, S. N. Desai, A. P. Singh, I. Y. Kim, C. W. Kim, S.-N. Park, A randomized, non-inferiority trial comparing two bivalent killed, whole cell, oral cholera vaccines (Euvichol vs Shanchol) in the Philippines. *Vaccine*. **33**, 6360–6365 (2015).
9. L. Qin, P. B. Gilbert, L. Corey, M. Juliana McElrath, S. G. Self, A Framework for Assessing Immunological Correlates of Protection in Vaccine Trials. *The Journal of Infectious Diseases*. **196** (2007), pp. 1304–1312.
10. S. Bolotin, S. L. Hughes, N. Gul, S. Khan, P. A. Rota, A. Severini, S. Hahné, A. Tricco, W. J. Moss, W. Orenstein, N. Turner, D. Durrheim, J. M. Heffernan, N. Crowcroft, What Is the Evidence to Support a Correlate of Protection for Measles? A Systematic Review. *J. Infect. Dis.* **221**, 1576–1583 (2020).
11. D. S. Khoury, D. Cromer, A. Reynaldi, T. E. Schlub, A. K. Wheatley, J. A. Juno, K. Subbarao, S. J. Kent, J. A. Triccas, M. P. Davenport, Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nat. Med.* **27**, 1205–1211 (2021).
12. P. B. Gilbert, D. C. Montefiori, A. B. McDermott, Y. Fong, D. Benkeser, W. Deng, H. Zhou, C. R. Houchens, K. Martins, L. Jayashankar, F. Castellino, B. Flach, B. C. Lin, S. O’Connell, C. McDanal, A. Eaton, M. Sarzotti-

Kelsoe, Y. Lu, C. Yu, B. Borate, L. W. P. van der Laan, N. S. Hejazi, C. Huynh, J. Miller, H. M. El Sahly, L. R. Baden, M. Baron, L. De La Cruz, C. Gay, S. Kalams, C. F. Kelley, M. P. Andrasik, J. G. Kublin, L. Corey, K. M. Neuzil, L. N. Carpp, R. Pajon, D. Follmann, R. O. Donis, R. A. Koup, Immune Assays Team§, Moderna, Inc. Team§, Coronavirus Vaccine Prevention Network (CoVPN)/Coronavirus Efficacy (COVE) Team§, United States Government (USG)/CoVPN Biostatistics Team§, Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. *Science*. **375**, 43–50 (2022).

13. D. Saha, R. C. LaRocque, A. I. Khan, J. B. Harris, Y. A. Begum, S. M. Akramuzzaman, A. S. G. Faruque, E. T. Ryan, F. Qadri, S. B. Calderwood, Incomplete correlation of serum vibriocidal antibody titer with protection from *Vibrio cholerae* infection in urban Bangladesh. *J. Infect. Dis.* **189**, 2318–2322 (2004).

14. K. Wiens, B. Jauregui, B. Arnold, K. Banke, D. Wade, K. Hayford, A. Costero-Saint Denis, R. Hall, H. Salje, I. Rodriguez-Barraquer, A. Azman, G. Vernet, D. Leung, Precision public health through serological biomarkers: An integrated surveillance platform to inform public health interventions. *Preprints* (2021), , doi:10.20944/preprints202110.0444.v1.

15. C. J. E. Metcalf, J. Farrar, F. T. Cutts, N. E. Basta, A. L. Graham, J. Lessler, N. M. Ferguson, D. S. Burke, B. T. Grenfell, Use of serological surveys to generate key insights into the changing global landscape of infectious disease. *Lancet*. **388**, 728–730 (2016).

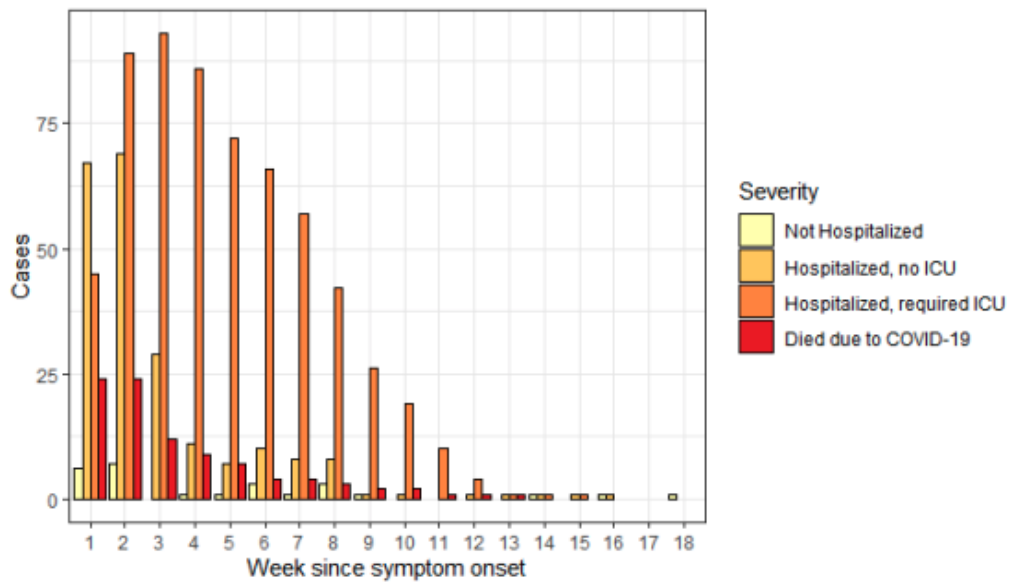
16. B. F. Arnold, H. M. Scobie, J. W. Priest, P. J. Lammie, Integrated Serologic Surveillance of Population Immunity and Disease Transmission. *Emerg. Infect. Dis.* **24**, 1188–1194 (2018).

## APPENDICES

### Appendix A: Supplement to Persistence and decay of human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in COVID-19 patients

**Figure A.1. Number of PCR positive cases with a sample taken during each week since symptom onset.**

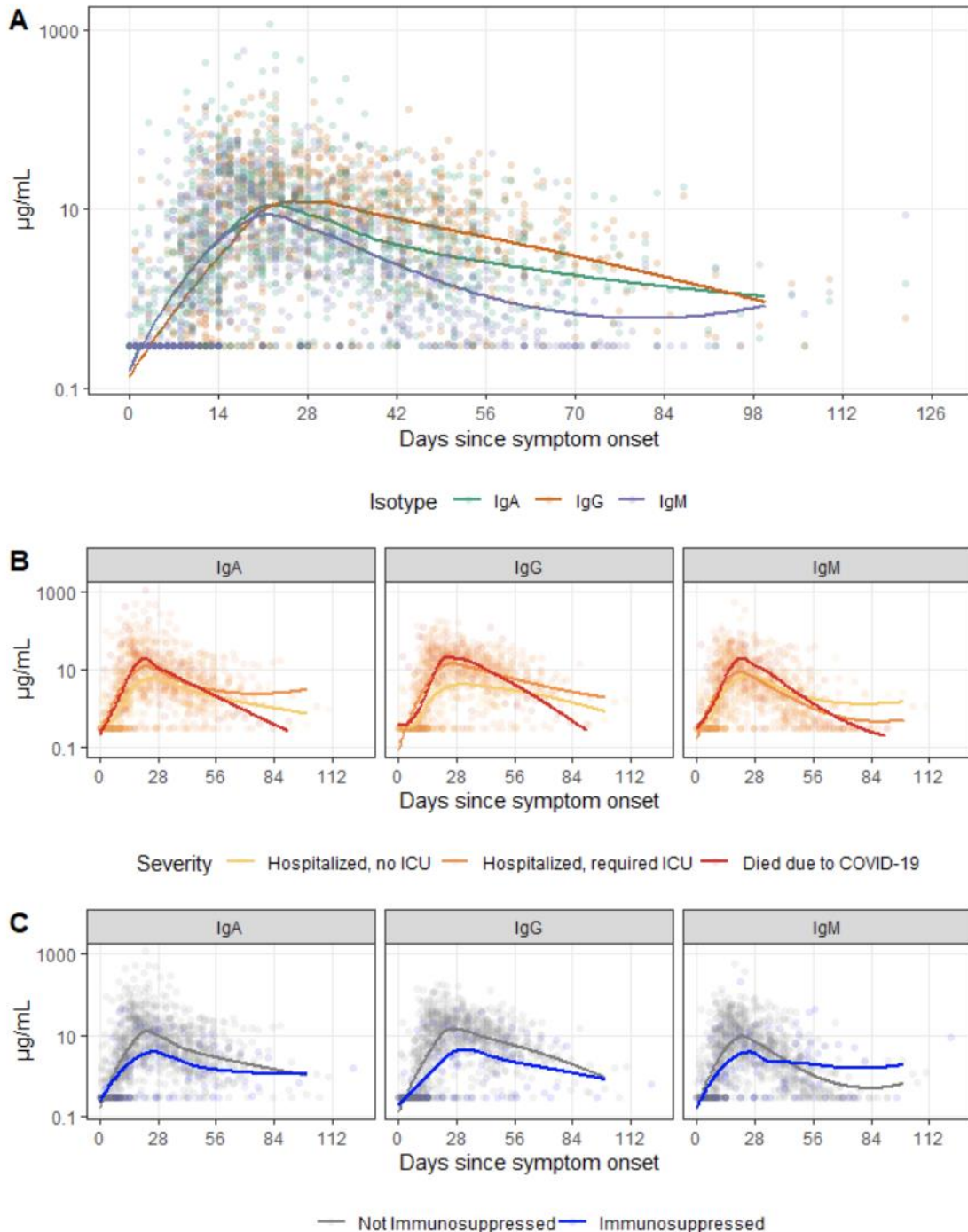
The date of symptom onset could not be determined for three individuals and the severity index was missing for one individual.





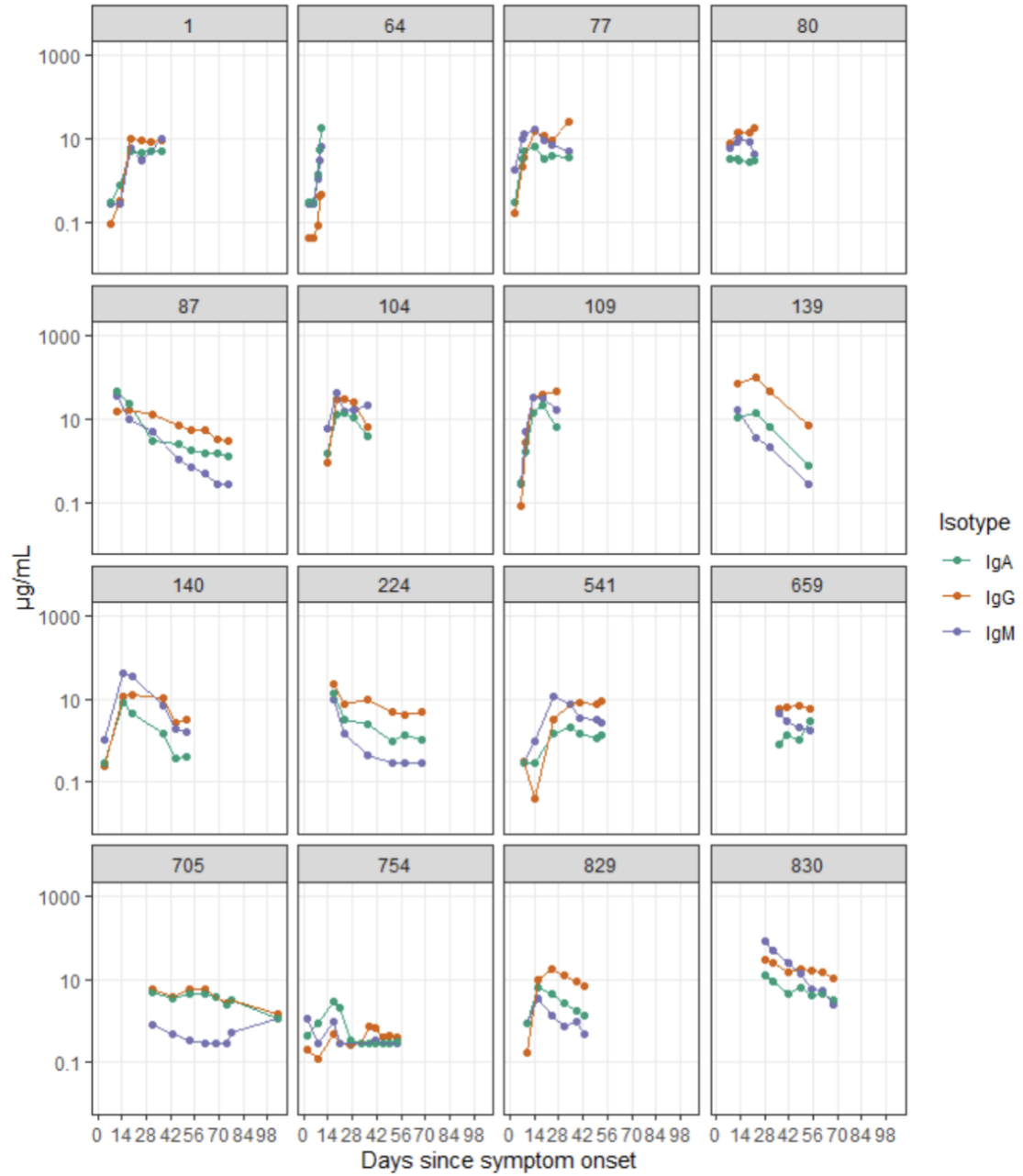
**Figure A.2. Smooth average measurements of IgG, IgM, and IgA against SARS-CoV-2 spike protein receptor binding domain among PCR positive cases across time.**

Limit of detection was artificially set at 0.3  $\mu\text{g/mL}$  for IgM and IgG to match that of IgA. Points were jittered horizontally. A) All cases are shown. Cases are categorized by B) clinical severity and C) immunosuppression status.



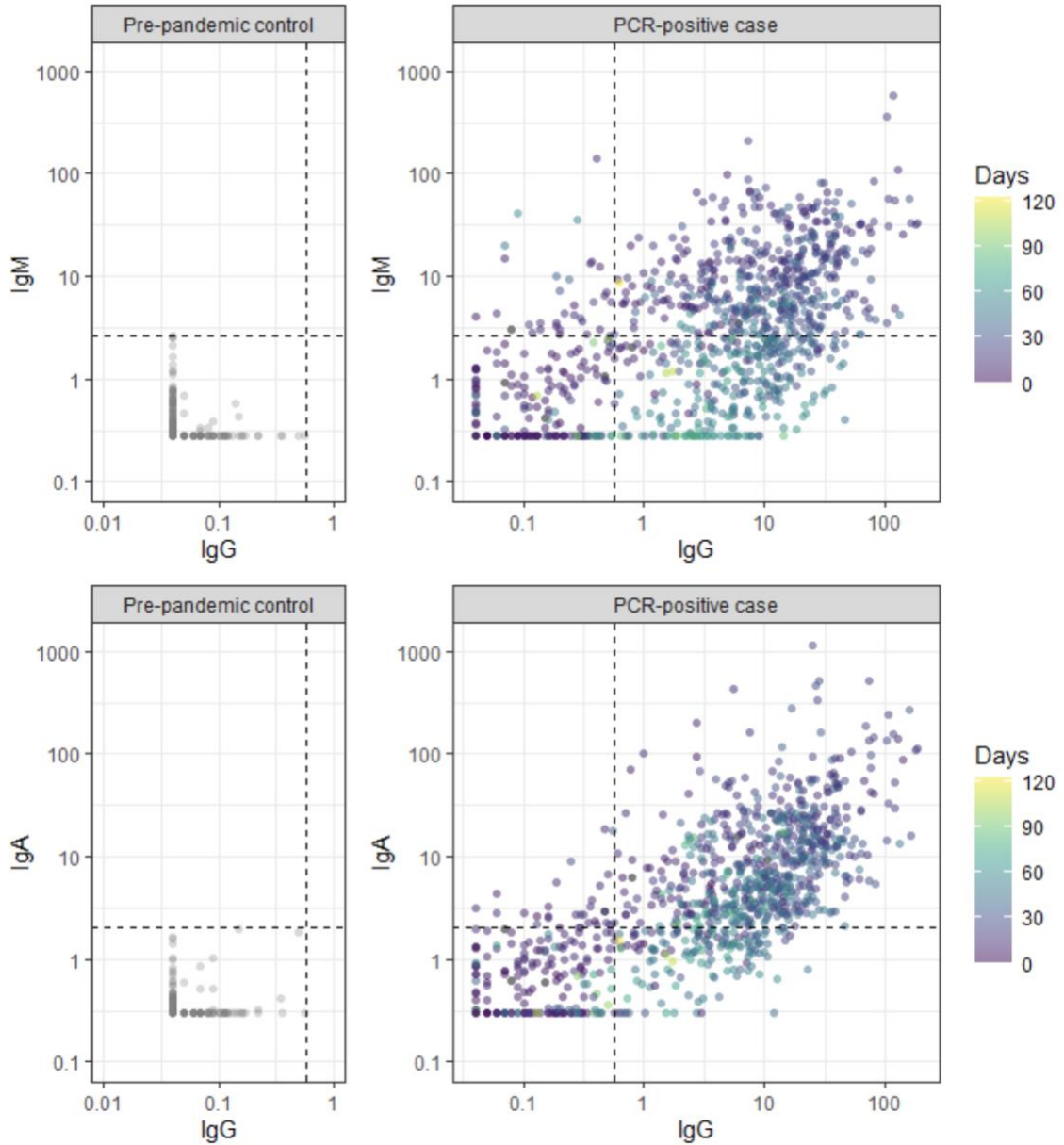
**Figure A.3. Individual trajectories for 16 randomly selected individuals with 4 or more measurements.**

Patient ID numbers are shown in gray boxes.



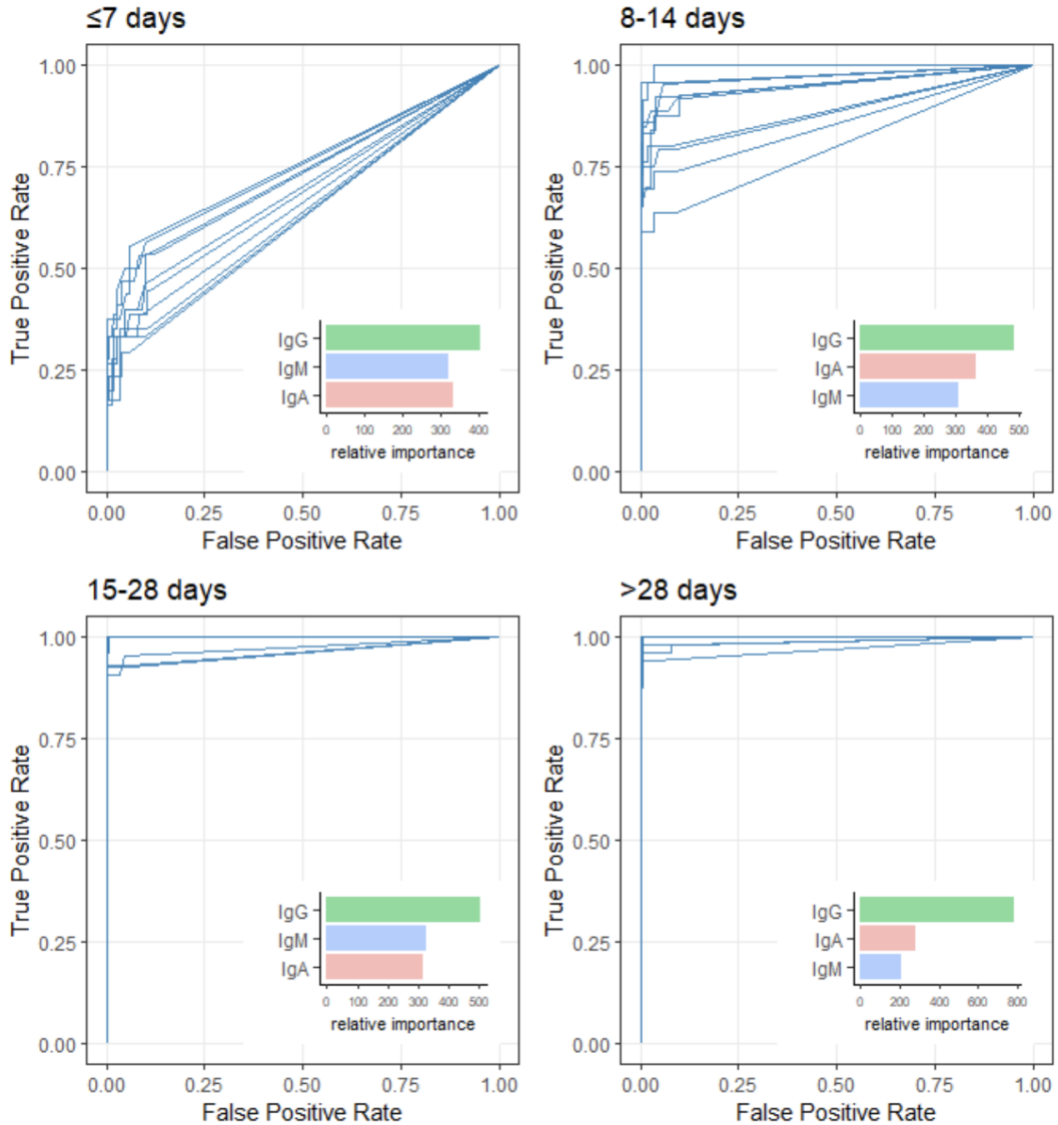
**Figure A.4. Measurements of IgG, IgM, and IgA against SARS-CoV-2 spike protein receptor binding domain among pre-pandemic controls and symptomatic PCR positive cases.**

Black dashed line is at 0.57  $\mu\text{g/mL}$  for IgG, 2.63  $\mu\text{g/mL}$  for IgM, and 2.02  $\mu\text{g/mL}$  for IgA.

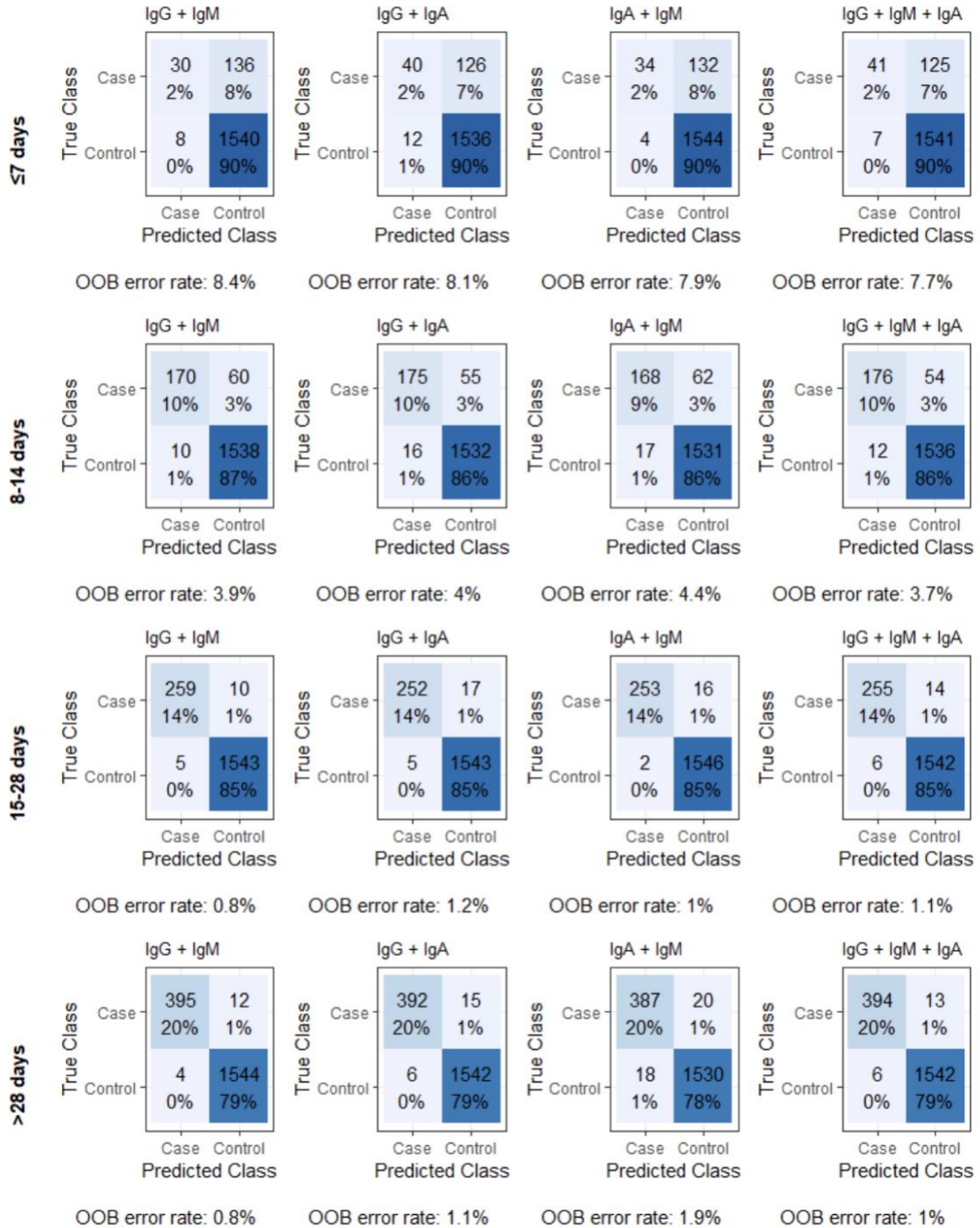


**Figure A.5. Receiver operating characteristic curve from random forest models and isotype contributions.**

Each panel shows the ROC curves for cross-validated random forest models fit to serological measurements taken (A) under 7 days (cvAUC: 0.64), (B) 8-14 days (cvAUC: 0.92), (C) 15-28 days (cvAUC: 1.00) and over 28 days (cvAUC: 1.00) after symptom onset of PCR positive cases and pre-pandemic controls. Each blue line is one of ten cross-validated ROC curves for a specific time point. Median relative importance of each serological marker is shown in each bar graph.

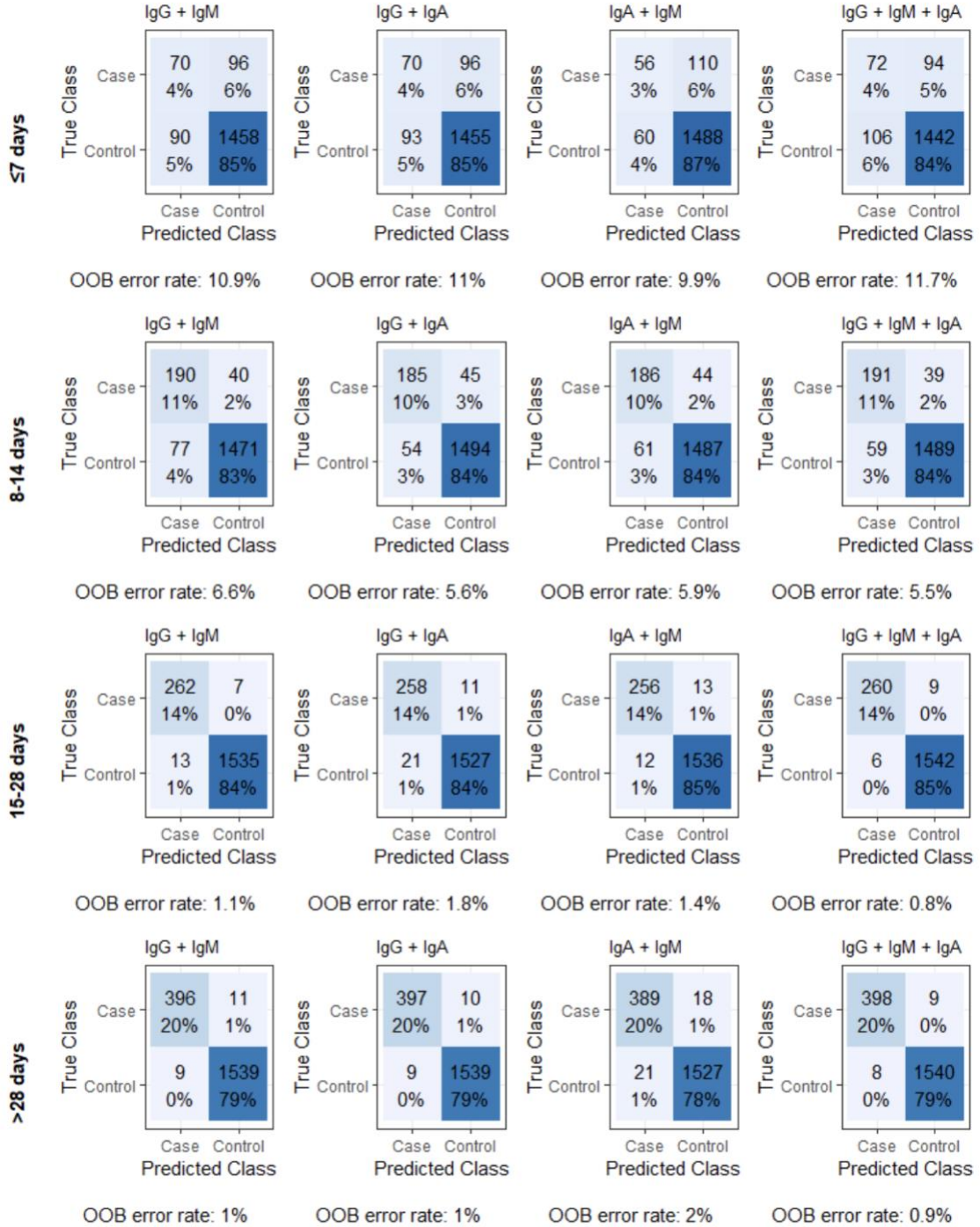


**Figure A.6. Confusion matrices and out-of-bag error estimates for random forest models.**



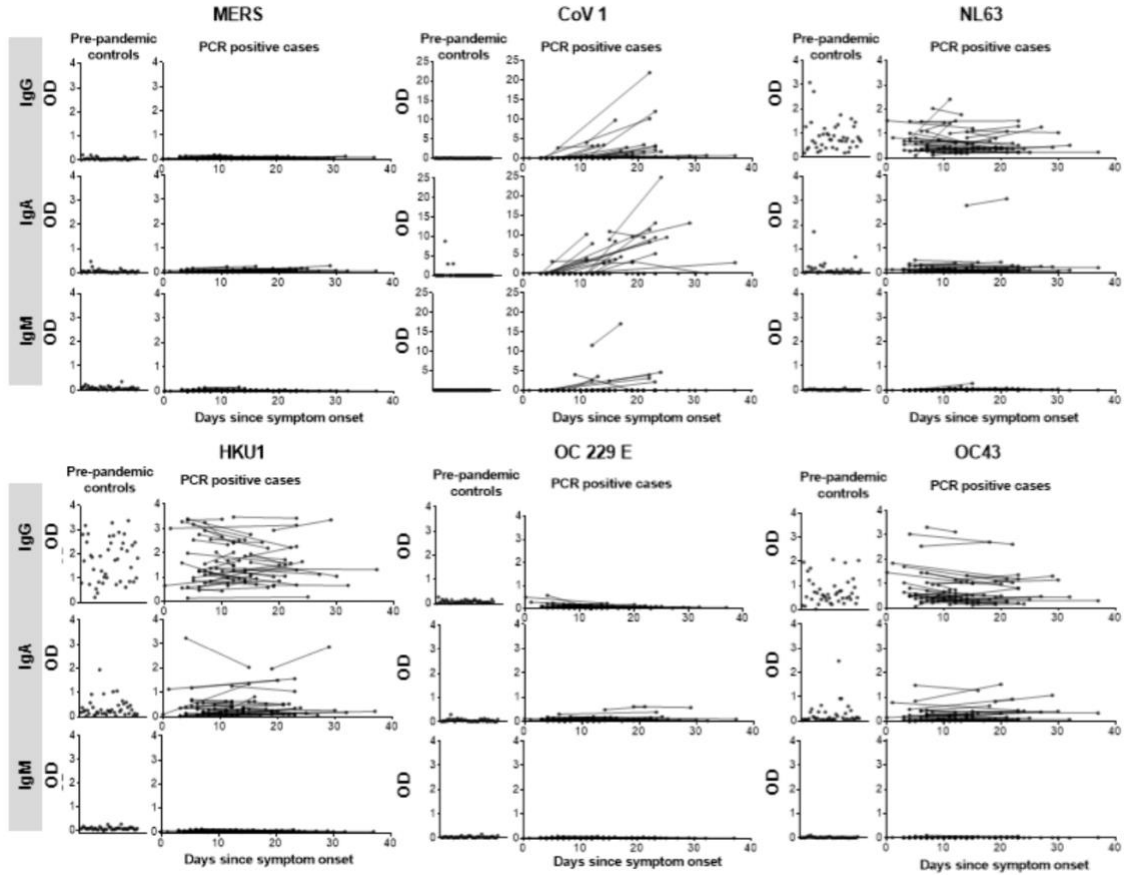
**Figure A.7. Confusion matrices and out-of-bag error estimates for random forest models with downsampled controls.**

Controls were downsampled to have the same number of samples as cases for a given period.



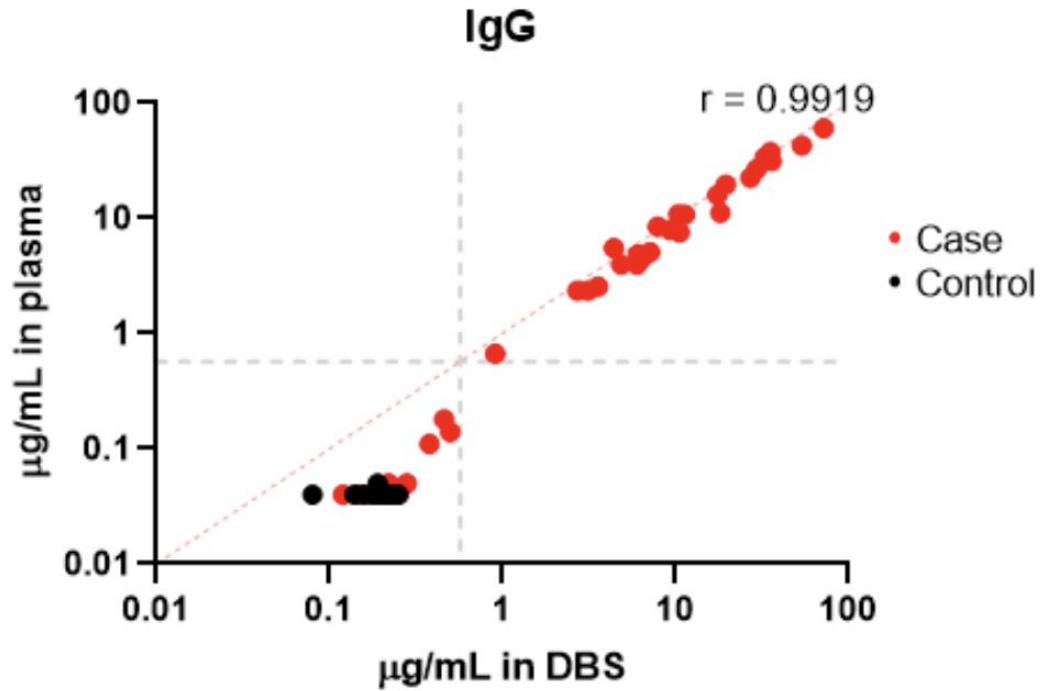
**Figure A.8. Measurements of IgG, IgA, and IgM against the RBD of other coronaviruses among pre-pandemic controls and PCR positive cases.**

Each dot represents a unique measurement of a serological marker (Row A: IgG, Row B: IgA, Row C: IgM) in pre-pandemic controls (left panels) and PCR positive cases (right panels) for each coronavirus. Each line connects measurements (dots) for individuals.



**Figure A.9. Correlation between plasma and dried blood spot measurements (DBS).**

Plot of anti-RBD antibody IgG measurement in plasma versus DBS of 20 COVID cases (at 2 timepoints) and 20 pre-pandemic controls. The Pearson correlation coefficient ( $r$ ) is shown. The dotted gray lines represent the concentration cut-off for seropositivity with plasma.





**Table A.1. Full amino acid sequences of the coronavirus receptor-binding domains (RBDs) used in this study.**

In parentheses are the GenBank accession numbers from which the RBDs derive. Underlined is the HRV-3C protease site, the 8xHis and streptavidin-binding peptide (SBP) purification tags.

Coronavirus	RBD Sequence
SARS-CoV-2	>SARS2_RBD (MN975262) RVQPTEIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGNGYNYLYRFRKSNLKPFFERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKGAGSSLEVLFQGPVGGSSHHHHHHHHGGSGSSMDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP
SARS-CoV-1	>SARS1_RBD (AAP13441.1) RVVPSGDVVRFPNITNLCPFGEVFNATRFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLN DLCSNVYADSFVVKGDVDRQIAPGQTVGIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRH GKLRPFERDISNVFSPDGKPCPPALNCYWPLNDYGFYTTTGIGYQPYRVVLSFELLNAPATVCGPKL GAGSSLEVLFQGPVGGSSHHHHHHHHGGSGSSMDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP
MERS	>MERS_RBD (AFY13307.1) EAKPSGSVVEQAEGVECDFSPLLSGTTPQVYNFKRLVFTNCNYNLTKLLSLFSVNDFTCSQISPAAIASNCYSSILILDYFSYPLSMKSDLVSSAGPISQFNKYQSFNSPTCLILATVPHNLTTITKPLKYSYINKCSRFLSDDRTEVPQLVNAVQYSPCVSIVPSTVWEDGDYRKLSPLEGGGWLVASGSTVAMTEQLQMGFGITVQYGTDTNSVCPKLEFANDTKIASQLGAGSSLEVLFQGPVGGSSHHHHHHHHGGSGSSMDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP
HKU1	>HKU1_RBD (AAT98580.1) TVKPVATVHRRIPDLPDCDIDKWLNNFNVPSPLNWERKIFSNCFNLSTLLRLVHTDSFSCNNFDESKIYGSCFKSIVLDKFAIPNSRRSDLQLGSSGFLQSSNYKIDTSSSCQLYSLPAINVTINNYNPSWNRRYGFNNFNLSHSHSVVYSRYCFVNNFTFCPCAKPSFASSCKSHKPPSASCPIGTNYRSCESTTVLDHTDWCRCSCLPDPITAYDPRSCSQKSLVGVGEHCAGFGVDEEKCGVLDGSSYNVSLCSTDAFLGWSYDTCVSNRNCNIFSNFILNGINSGTTCSDNLLQPNTEVFTDVCVDYDLYGITGQGI FKEVSAVYYNSWQNLLYDSNGNIIGFKDFVTNKTYNIFPCYAGGAGSSLEVLFQGPVGGSSHHHHHHHHGGSGSSMDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP
OC229E	>OC229E_RBD (AAK32191) VSLPVYHKHMFIVLYVDFKQSGGGKCFNCYPAGVNI TLANFNETKGPLCVDTSHTTTKYVAVYANVGRWSASINTGNCPFSFGKVNNFVKFGSVCFLKDI PGGCAMP IVANWAYS KYTYITGLYVSWSDGDGITGVPQ PVEGVGAGSSLEVLFQGPVGGSSHHHHHHHHGGSGSSMDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP
OC43	>OC43_RBD (AAT84362) RRKPNLPNCNIEAWLNDKSVPSPLNWERKTFSNCFNMSSLMSFIQADSFTCNNIDAAKIYGMCFSSITIDKFAIPNGRKVDLQLGNLGYLQSFNYRIDTTATSCQLYYNLPAANVSVSRFPNTPWNKRFGFIEDSVFKRPPAGVLTNHDDVVAQHCFKAPKNFCPCCKLNGSCVGGSGPKNNGIGTFCPAGTNYLTCNDLCTPDPIFTFTGT YKCPQTKSLVGI GEHC SGLAVKSDYCGGNSCTCRPQAFLGWSADSLQGDKNIFANF I LHDVNSGLTCS TDLQKANTGAGSSLEVLFQGPVGGSSHHHHHHHHGGSGSSMDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP
NL63	>NL63_RBD (AKT07952) QHTDINFATASFGGSCYVCKPHQVNI SLNGNTSVCVRTSHFSIRYIYNRVKSQSPGDSSWHIYLKSGTCCPFSFKLNNFQKFKTICFSTVEVPGSCNFPLEATWHYTSYTI V GALYVTWSEGNSTGVPYVPSGI GAGSSLEVLFQGPVGGSSHHHHHHHHGGSGSSMDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP

**Table A.2. Predictive accuracy of multiple isotypes for classifying controls and cases over time since symptom onset.**

Random forest models were used to calculate cvAUC. The isotype cut-offs chosen for calculating sensitivity were the maximum concentration ( $\mu\text{g/mL}$ ) found among pre-pandemic controls (IgG: 0.57, IgM: 2.63, IgA: 2.02). Samples with measurements above at least one cut-off were classified as cases.

Isotypes	Days since symptom onset	cvAUC (95% CI)	Sensitivity (95% CI)
IgA + IgG	$\leq 7$ days	0.68 (0.59–0.78)	0.10 (0.06–0.15)
	8-14 days	0.91 (0.87–0.96)	0.58 (0.51–0.66)
	15-28 days	0.99 (0.96–1.00)	0.96 (0.93–0.99)
	>28 days	0.99 (0.97–1.00)	0.95 (0.91–0.98)
IgM + IgG	$\leq 7$ days	0.68 (0.59–0.77)	0.09 (0.04–0.14)
	8-14 days	0.92 (0.87–0.96)	0.59 (0.51–0.66)
	15-28 days	0.99 (0.97–1.00)	0.96 (0.93–0.99)
	>28 days	0.99 (0.98–1.00)	0.96 (0.92–0.99)
IgM + IgA	$\leq 7$ days	0.65 (0.56–0.75)	0.09 (0.05–0.15)
	8-14 days	0.90 (0.85–0.95)	0.59 (0.52–0.66)
	15-28 days	0.98 (0.95–1.00)	0.95 (0.92–0.98)
	>28 days	0.98 (0.95–1.00)	0.74 (0.66–0.81)
IgM + IgA + IgG	$\leq 7$ days	0.69 (0.60–0.78)	0.11 (0.06–0.16)
	8-14 days	0.92 (0.88–0.97)	0.63 (0.56–0.69)
	15-28 days	0.99 (0.97–1.00)	0.97 (0.95–0.99)
	>28 days	0.99 (0.98–1.00)	0.96 (0.92–0.99)

**Table A.3. Parametric estimates of median time to seroconversion for each isotype by different patient characteristics.**

The isotype cutoffs chosen for seroconversion were the maximum concentration ( $\mu\text{g/mL}$ ) found among pre-pandemic controls (IgG: 0.57, IgM: 2.63, IgA: 2.02). All models assumed that time to event followed a Weibull distribution. Bootstrap 95% confidence intervals are shown in parentheses. Not Hospitalized and Hospitalized, no ICU groups were combined due to small sample size.

Isotype	Characteristic	50th percentile (95% CI)	Difference (95% CI)
IgA	Age		
	<65 years	11.8 (10.5 – 13.1)	
	$\geq$ 65 years	13.7 (10.6 – 17.5)	1.9 (-0.9 – 5.4)
	Sex		
	Male	11.8 (10.5 – 13.3)	
	Female	16.4 (12.4 – 21.8)	4.6 (1.1 – 9.3)
	Severity		
	Not Hospitalized / Hospitalized, no ICU	12.2 (10.9 – 13.6)	
	Hospitalized, required ICU	7.3 (5.5 – 9.5)	-4.9 (-6.9 – -2.8)
Died due to COVID-19	9.0 (6.2 – 12.2)	-3.2 (-6.1 – 0.2)	
IgG	Age		
	<65 years	10.7 (9.6 – 11.9)	
	$\geq$ 65 years	13.3 (10.4 – 16.8)	2.6 (-0.0 – 5.9)
	Sex		
	Male	10.7 (9.7 – 11.8)	
	Female	14.2 (11.0 – 18.2)	3.5 (0.7 – 7.1)
	Severity		
	Not Hospitalized / Hospitalized, no ICU	10.9 (9.6 – 12.2)	
	Hospitalized, required ICU	6.9 (5.2 – 8.7)	-4.0 (-5.7 – -2.2)
Died due to COVID-19	10.1 (6.8 – 14.2)	-0.8 (-4.1 – 3.1)	
IgM	Age		
	<65 years	12.0 (10.7 – 13.6)	
	$\geq$ 65 years	13.8 (9.9 – 18.8)	1.8 (-1.5 – 6.2)
	Sex		
	Male	12.1 (10.7 – 13.7)	
	Female	17.8 (12.9 – 23.7)	5.7 (1.4 – 11.0)
	Severity		
	Not Hospitalized / Hospitalized, no ICU	12.3 (10.7 – 14.2)	
	Hospitalized, required ICU	7.9 (5.8 – 10.2)	-4.4 (-6.9 – -2.0)
Died due to COVID-19	7.3 (4.9 – 10.8)	-5.0 (-7.8 – -1.4)	

## **Appendix B: Supplement to Identifying recent cholera infections using a multiplex bead serological assay**

### **Method B.1. Overview of sample selection**

#### *Prediction-based sampling*

We aimed to select a sample of the original cohort data from Bangladesh to test out for this analysis of a multiplex bead assay. Rather than attempt to handpick individuals using descriptive guidelines, we decided to choose the sample that best predicts the rest of the cohort. The sample with the best prediction accuracy should have attributes that roughly reflect that of the whole cohort, be they demographic (age, sex, blood type) or test-based (vibriocidal, ELISAs).

The process for making these predictions was as follows:

- Choose a random sample of individuals to create the “training set”, for a given scenario (explained below)
- Fit a random forest model to those individuals to classify individuals as recently infected or not
- Create a “balanced test set” (explained below)
- Use the random forest to predict the outcomes in the balanced test set
- Evaluate the predictions with cross-validated area under the ROC curve (cvAUC)

We looked at three scenarios for selecting the training set, all of which included 20 individuals who are under 10 (‘children’) and 20 individuals who are 10 and older (‘adults’). The scenarios only differ by the number of household contacts used. In the first scenario, we randomly draw 10 adults who were index cases and 10 adults who were household contacts. In the second scenario, we draw only index cases and no household contacts. In the third scenario, we draw from the pool of index and household contacts indiscriminately; letting the predictive ability of the models choose the number of household contacts.

After selecting the training set and fitting the random forest model, we predicted a “balanced test set”. The test set consisted of the observations that were not chosen in the training set. We took a large random sample with replacement of the test set, but balanced such that (a) there were as many seropositive (as predicted by the model) observations as seronegative observations and (b) the seropositive observations were evenly distributed over their time since infection. This is important because the original cohort was skewed towards having more seropositive observations with lower times since infection. Having an equal number of seropositive and seronegative cases ensured that we did not reward models for erring on the side of making positive or negative predictions. By sampling across time since infection, we hoped to improve the model performance at times further from infection; the serosurvey model performed quite well at short time intervals but poorly at longer horizons.

After resampling 10,000 times, we chose a set that had the highest cvAUC for the third scenario. This included 38 cases and 2 household contacts.

#### Sample inventory

After investigation of freezer stocks, samples from 39 of 40 individuals were found. Samples from two additional cases were added to the set, giving a selection of 245 samples from 41 individuals (39 cases and 2 household contacts).

#### Additional sample selection

After initial testing was completed, we decided to supplement the dataset with samples from 10 individuals (9 cases and 1 contact) in order to balance out the ages of individuals. In particular, we were concerned about the lack of individuals with samples between the ages of 10 and 18 in the original sample. To limit the influence of boosted antibody responses from reinfection/exposure during the follow-up period, we removed any data points that were part of or after a greater than a 2 fold-rise between measurements in vibriocidal Ogawa titers >90 days post initial infection. Our final sample included 305 samples from 51 individuals (48 cases and 3 contacts).

## Method B.2. Procedure for calculating the relative antibody unit from median fluorescence intensity measurements

On each 384-well plate, a dilution series of positive control wells was included to adjust for between-plate variability. We used this to calculate the relative antibody unit (RAU) for all samples on the plate. The RAU is the expected dilution of the positive control sample needed to get the same MFI as the sample.

### Dilution series

Serum samples from 5 patients with culture-confirmed *V. cholerae* O1 collected 7 days after symptoms were combined to create a serum pool. Of the total 19 plates run, 7 had four dilution points (1:100, 1:400, 1:1,600, 1:6,400). After deciding to expand the range of dilutions, the next 12 plates had eight dilution points (1:10, 1:40, 1:160, 1:640, 1:2,560, 1:10,240, 1:40,960, 1:163,840). Additionally, blank control wells (i.e. not including any sera) were run on every plate. All samples were run in triplicate and MFI values were averaged.

### Model parameterization and fit

For each antigen, we fit four-parameter log-logistic models to each plate's dilution series. As previously described (1), the relationship between dilution and MFI is defined as follows:

$$\begin{aligned} Y_i &= \log(\text{median fluorescence intensity}) \text{ for sample } i \\ x_i &= \text{dilution for sample } i \\ Y_i &= c + \frac{d - c}{1 + \exp(b \times (\log(x_i) - \log(e)))} \end{aligned}$$

We chose this parameterization given its frequent use for modeling dose-response relationships and that we only had four dilution points for many of our plates. For plates with seven dilution points, we used the `drc` package in R (1). For plates with four dilution points, we decided to implement a Bayesian framework so that we could fit the model with prior distributions informed by the seven dilution series wells:

$$\begin{aligned} b &\sim \text{Normal}(\mu_b, 1) \\ c &\sim \text{Normal}(\mu_c, 0.5) \\ d &\sim \text{Normal}(\mu_d, 20) \\ e &\sim \text{Normal}(\mu_e, 100) \end{aligned}$$

Specifically,  $\mu_c$ , the mean of the prior distribution for  $c$  (the lower bound of the logistic curve), was set equal to the log(MFI) for the blank sample for each plate. The variance parameter of the prior distribution for  $c$  was set at 0.5 to be slightly larger than observed among blank samples. The mean values for the other priors ( $\mu_b, \mu_d$ , and  $\mu_e$ ) were set equal to the average values estimated from the seven dilution models using the drc package. Variance values for the prior distributions of  $b, d$ , and  $e$  were selected to be sufficiently large to be relatively uninformative.

#### Relative antibody unit calculation

For each antigen on each plate, the mean value for each parameter was used to calculate the relative antibody unit (RAU). To avoid extrapolation, all samples with a calculated RAU above 1:100 or below 1:100,000 were set equal to those values. Any samples with MFI values falling outside of the logistic curve were also set to the threshold value.

### Method B.3. Univariate decay model equations

We fit univariate decay models to describe the antibody dynamics of individuals for each biomarker. All models had two components: 1) a measurement model and 2) a decay model. This data set contains  $n$  individuals (indexed by  $i$ ), and  $m$  total measurements (indexed by  $j$ ). Each marker is modeled separately.

#### Measurement Models

We had several different types of measurements of biomarkers (multiplex bead assay (MBA) relative antibody unit (RAU), Vibriocidal titers, and ELISA concentrations) to fit models to. We assumed that the observed data were normally distributed ( $f_N$  denotes the normal density) around the true (unobserved) value. Each different model accounted for censored observations in various ways based on how the data was generated.

- $Y_{ij}$  = the true value at the time of the  $j$ th measurement for individual  $i$ .
- $Y_{ij}^*$  = the observed value at the time of the  $j$ th measurement for individual  $i$ .
- $\sigma$  = measurement error variance

#### *Multiplex bead assay relative antibody unit measurement model*

For our analysis, the RAU was inverted and log10 transformed. Values can only take a value above -5 and below -2. Any value set at -5 and -2 are censored values and are treated as so:

$$Pr(Y_{ij}^*|Y_{ij}, \sigma) = \begin{cases} \int_{-\infty}^{-5} f_N(Y_{ij}|\sigma) & \text{if } Y_{ij}^* = -5 \\ f_N(Y_{ij}|\sigma) & \text{if } -5 < Y_{ij}^* < -2 \\ \int_{-2}^{\infty} f_N(Y_{ij}|\sigma) & \text{if } Y_{ij}^* = -2 \end{cases}$$

#### *Vibriocidal assay measurement model*

Vibriocidal titers were divided by 5 and log2 transformed. The vibriocidal assay outputs a measurement of the highest dilution where the vibriocidal reaction still occurs. Therefore, the true dilution is between the reported dilution and the next dilution (i.e. interval censored).  $V_{max}$ , the largest measureable log titer, was 11.  $V_{min}$ , the smallest measureable log titer was 0.

$$Pr(Y_{ij}^*|Y_{ij}, \sigma) = \begin{cases} \int_{-\infty}^{V_{min}} f_N(Y_{ij}|\sigma) & \text{if } Y_{ij}^* = V_{min} \\ \int_{Y_{ij}}^{Y_{ij}+1} f_N(Y_{ij}|\sigma) & \text{if } V_{min} < Y_{ij}^* < V_{max} \\ \int_{V_{max}}^{\infty} f_N(Y_{ij}|\sigma) & \text{if } Y_{ij}^* = V_{max} \end{cases}$$



### *ELISA measurement model*

ELISA measurements were all log10 transformed. No values were at the limit of detection so censoring was not accounted for.

$$Pr(Y_{ij}^* | Y_{ij}, \sigma) = f_N(Y_{ij}, \sigma)$$

### *Kinetic Models*

We also explored different modes of decay (exponential vs. biphasic). Additionally, we also investigated differences in decay between individuals with different of demographic characteristics.

The following variables and parameters are shared across all kinetic models:

- $T_{ij}$  = time of sample collection post-infection for individual  $i$  and measurement  $j$ .
- $\omega_i$  = baseline value for individual  $i$
- $\lambda_i$  = boost for individual  $i$  (restricted to values greater than 0)
- $D$  = time between infection and initial rise (set at 5 days)
- $\mu^\omega$  = average individual's baseline rate
- $\mu^\lambda$  = average individual's boost from baseline at day  $D$  (restricted to values greater than 0)
- $\Sigma$  = covariance matrix for average baseline and boost

All models follow the same general temporal pattern:

- Individuals start at their baseline values  $\omega_i$
- After  $D$  days, an individuals value immediately increases by  $\lambda_i$
- Over time, an individuals value decays

We allow for baseline values and boosts to vary between individuals, but assume the decay rate is shared across individuals.

### *Exponential decay model*

For the exponential model, we assume that decay follows an exponential pattern with the decay parameter  $\delta$  (restricted to be greater than zero).

$$Y_{ij} = \begin{cases} \omega_i & T_{ij} < D \\ \omega_i + \lambda_i \times e^{-\delta(T_{ij}-D)} & T_{ij} \geq D \end{cases}$$
$$\begin{pmatrix} \omega_i \\ \lambda_i \end{pmatrix} \sim \text{MVN} \left( \begin{pmatrix} \mu^\omega \\ \mu^\lambda \end{pmatrix}, \Sigma \right)$$

### *Biphasic decay model*

For the biphasic decay model, we assume that each marker's value is determined by two independent components that each decay exponentially. The proportion of these two components is unknown.

- $\theta_1$  = decay rate for first component (restricted to be greater than zero)

- $\theta_2$  = decay rate for second component (restricted to be greater than zero)
- $\alpha$  = proportion of boost due to first component (restricted between 0 and 1)

$$Y_{ij} = \begin{cases} \omega_i & T_{ij} < D \\ \omega_i + \lambda_i \left( \alpha e^{-\theta_1(T_{ij}-D)} + (1-\alpha)e^{-\theta_2(T_{ij}-D)} \right) & T_{ij} \geq D \end{cases}$$

$$\begin{pmatrix} \omega_i \\ \lambda_i \end{pmatrix} \sim \text{MVN} \left( \begin{pmatrix} \mu^\omega \\ \mu^\lambda \end{pmatrix}, \Sigma \right)$$

### *Exponential model including individual covariates (e.g. age group)*

To understand how kinetics varied by individual-level attributes, we included age (<10 vs 10+ years), sex (male vs female), blood group (O group vs. non-O group), and infecting serotype (Ogawa vs. Inaba) as binary variables in the model. These covariates were allowed to modify the initial baseline, boost and decay rate.

- $X_i$  = covariate of individual  $i$  (e.g. 0 if < 10 years and 1 if 10+years)
- $\beta^\omega$  = fixed effect of covariate on baseline
- $\beta^\lambda$  = fixed effect of covariate on boost
- $\beta^\delta$  = fixed effect of covariate on decay

$$\mu_{ij} = \begin{cases} \omega_i + \beta^\omega X_i & T_{ij} < D \\ \omega_i + \beta^\omega X_i + (\lambda_i + \beta^\lambda X_i) \times e^{-(\delta + \beta^\delta X_i)(T_{ij}-D)} & T_{ij} \geq D \end{cases}$$

$$\begin{pmatrix} \omega_i \\ \lambda_i \end{pmatrix} \sim \text{MVN} \left( \begin{pmatrix} \mu^\omega \\ \mu^\lambda \end{pmatrix}, \Sigma \right)$$

## Method B.4. Sample weighting

The samples used to train random forest classification models are heavily skewed towards times in the early acute and convalescent period due to the higher density of blood draws during that period and the limited loss to follow-up early on (**Figure B.2**). This skewed distribution of infection times likely does not represent the expected more uniform distribution of infection times in a study population during a cross-sectional survey. We would expect that the assessments of model fit might be misleading or overly optimistic if this skew is not accounted for.

We attempted to re-weight the samples used to train random forest models in each class based on an expected distribution of time since infection. Additionally, we used weighting to account for class imbalance such that the sum of weights among recently infected and non-recently infected were made equal.

### Expected distribution of infection times

For this analysis, we decided to assume a constant hazard of infection with an annual incidence rate of 10%. When conducting a cross-sectional serosurvey at one point in time, we would expect that the time since last infection should be exponentially distributed. This could be modified given a different understanding of distribution of last infection times. Using the cumulative distribution function of an exponential distribution, we can understand the expected proportion of samples from each time slice.

### Defining time slices

As shown in **Figure B.2**, the day of sample collection we have for each sample are clustered around certain time points. In order to properly reweight the samples we have, we need to consider what unobserved infections the samples are 'standing in' for.

Given a particular infection window (e.g. 200-day), we assigned each sample to whether they were inside or outside of the window. We assumed baseline samples (collected <5 days post infection) of cases and household contacts were always outside of the infection window.

Next, we further divided the time inside and outside the infection window into time slices. For cases, samples were collected around 2, 7, 30, 90, 180, 270, 360, 540, 720, 900 and 1080 days. We used these times to define the time slices using the average of consecutive time points as dividers. We assumed that samples from any case after 540 days as well as uninfected contacts belonged to the same time slice.

200-day Infection Window	Status	Day	Time Slice	n (%)
Inside	Case	7	[5,18.5)	46 (27)
		30	[18.5,60)	46 (27)
		90	[60,135)	42 (24)

200-day Infection Window	Status	Day	Time Slice	n (%)
		180	[135,200)	39 (23)
Outside	Case	2	[0,5)	48 (36)
		270	[200,315)	11 (8)
		360	[315,450)	12 (9)
		540	[450,Inf)	25 (19)
		720	[540,Inf)	1 (1)
		900	[540,Inf)	25 (19)
		1080	[540,Inf)	1 (1)
Outside	Contact	2	[540,Inf)	3 (2)
		7	[540,Inf)	3 (2)
		30	[540,Inf)	3 (2)

All samples belonging to the same time slice are equally weighted and stand in for all potential infection times within the slice. If the proportion of samples for given time slice does not match what is expected from an exponential distribution, they can now be properly weighted.

#### Weight calculation

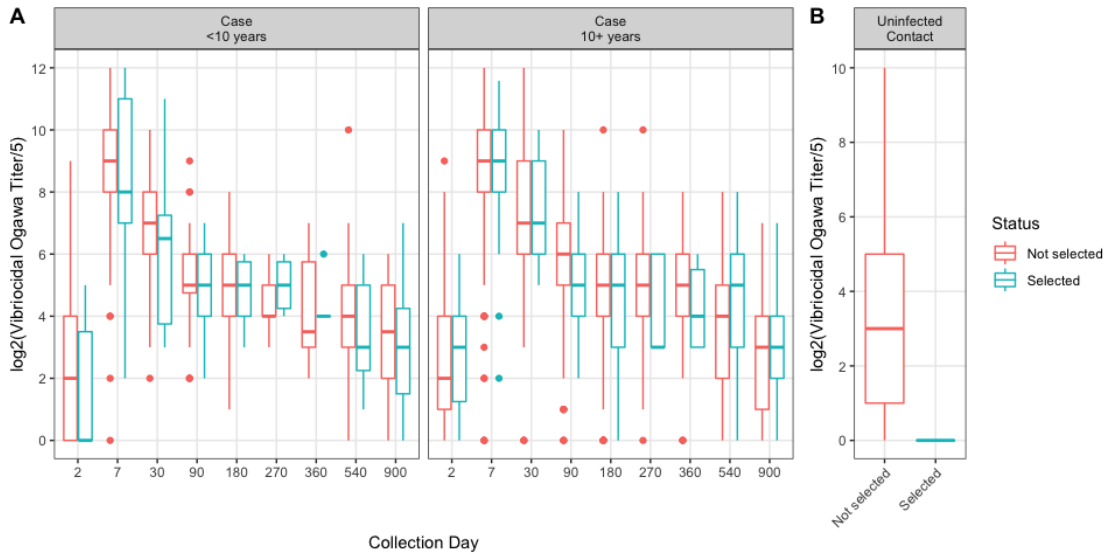
We calculated weights to account for both class imbalance and distribution of infection times. To account for class imbalance, we simply created weights so the sum of weights among samples inside and that of samples outside the window were equal. For the class weight, samples within each class were equally weighted.

To account for the distribution of infection times, we calculated both the observed and the expected proportion of samples to be in each time slice. The calculated time-based weight to account for infection time is the ratio of the expected versus observed proportion (**Figure B.7**).

The final weight used in our analyses is the product of the class weight and the time-based weight. It was recalculated for every different length of infection window and each fold in cross-validation.

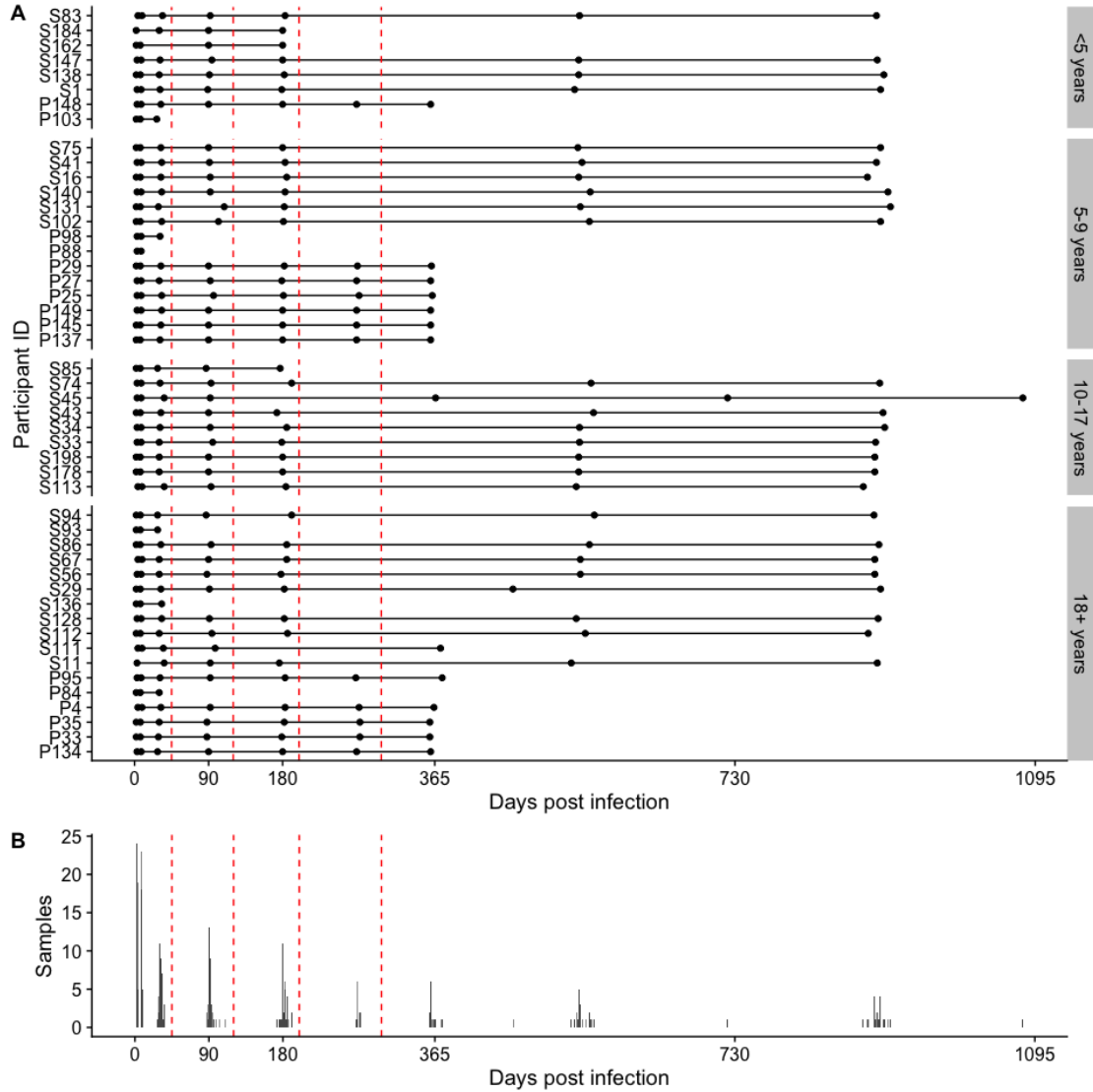
### Figure B.1. Vibriocidal Ogawa measurements among individuals selected from the SMIC and PIC cohorts

Vibriocidal titer measurements were transformed by dividing the measurement by 5 and then taking the logarithm (base 2) (A). X-axis indicates the approximate collection day for each measurement for cases. All measurements for uninfected household contacts are shown regardless of collection day (B).



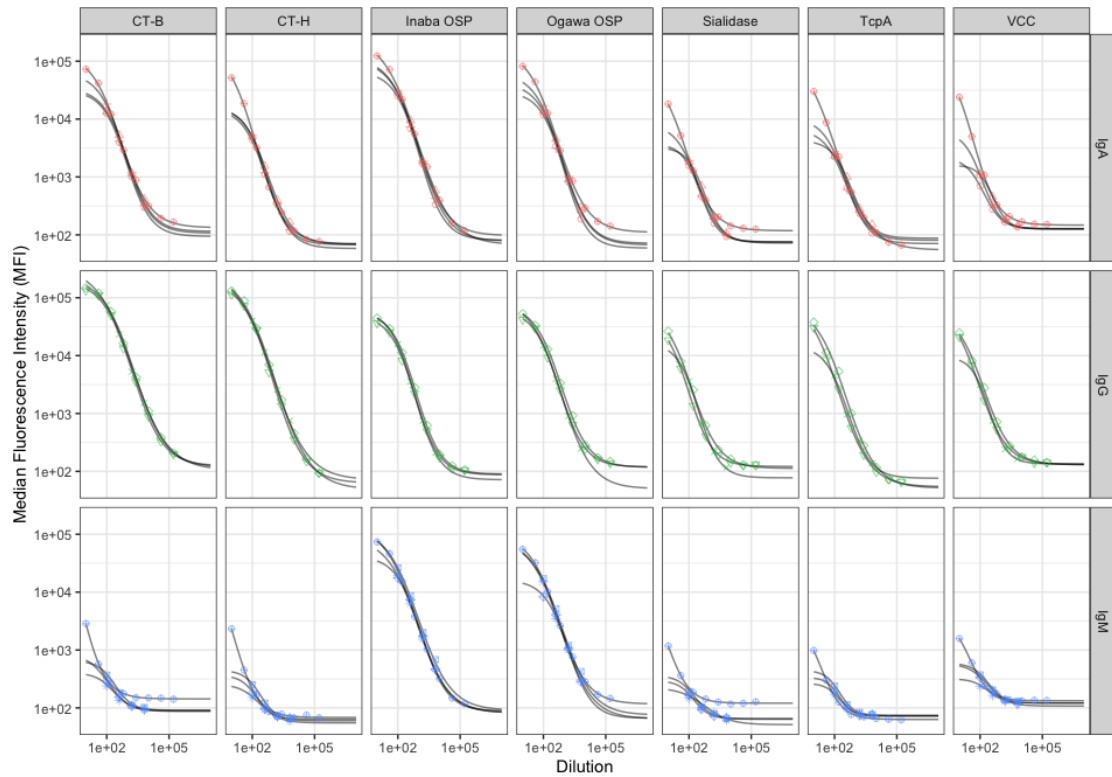
**Figure B.2. Timing of sample collection among culture confirmed cases relative to date of infection**

Each point represents a blood sample collection occurred for a confirmed case (A). The daily number of samples is shown on the y-axis. Red dashed lines are placed at the threshold values for infections windows (45, 120, 200, and 300 days) (B).



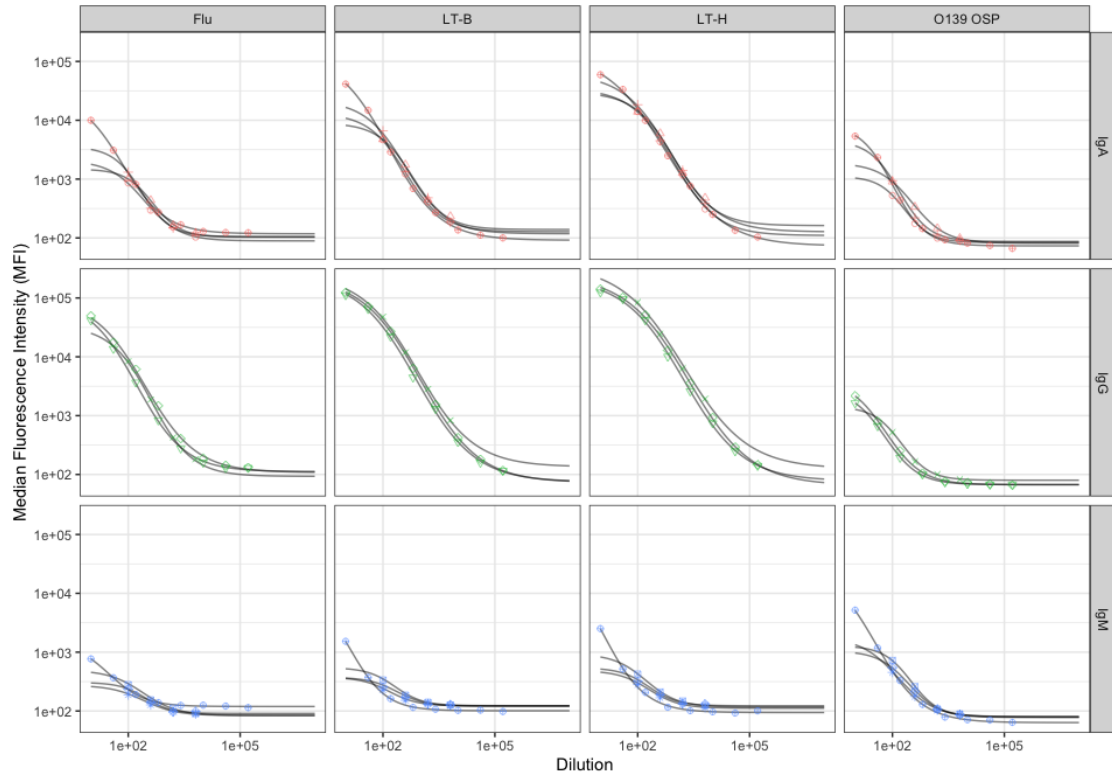
**Figure B.3. Postive control dilution curve and four parameter log-logistic curve fits for *Vibrio cholerae* O1 antigen multiplex bead assay markers**

Each point represents the median fluorescence intensity measurement (y-value) for a dilution (x-value) of pooled convalescent sera from confirmed positive *Vibrio cholerae* O1 patients. Each plate's dilution series were fit using a four-parameter log-logistic regression (black lines). Shape of each point is unique to each plate.



**Figure B.4. Postive control dilution curve and four parameter log-logistic curve fits for additional antigen multiplex bead assay markers**

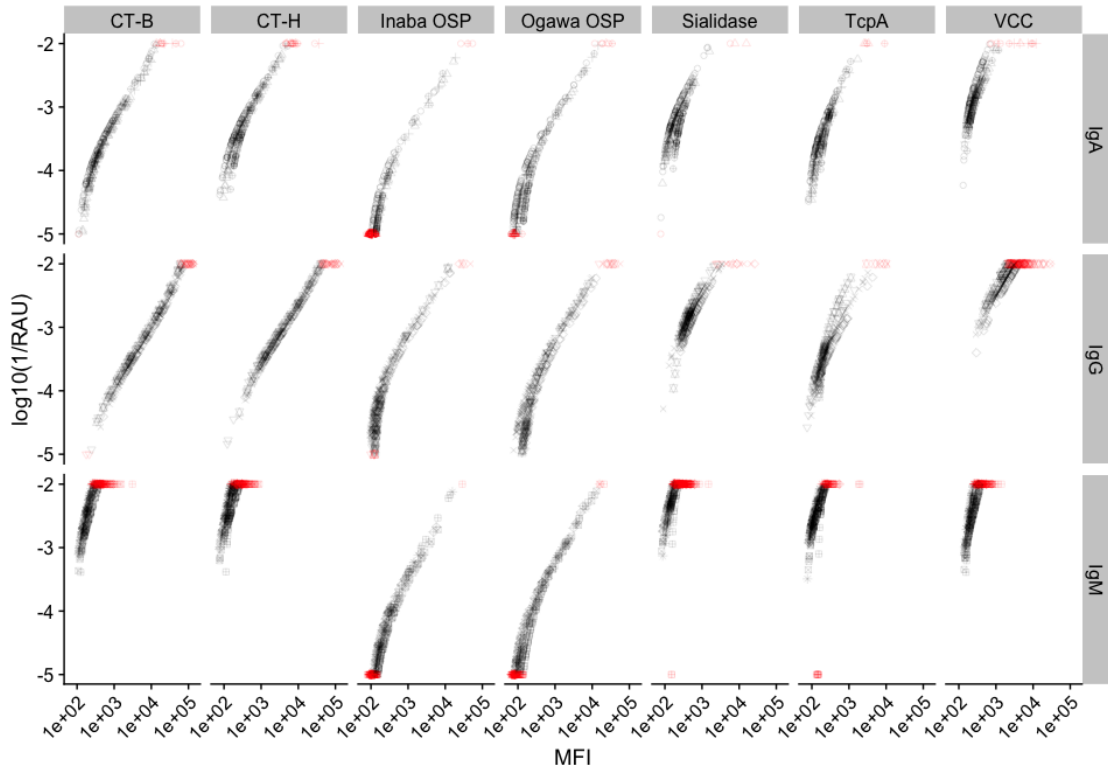
Each point represents the median fluorescence intensity measurement (y-value) for a dilution (x-value) of pooled convalescent sera from confirmed positive *Vibrio cholerae* O1 patients. Each plate's dilution series were fit using a four-parameter log-logistic regression (black lines). Shape of each point is unique to each plate.





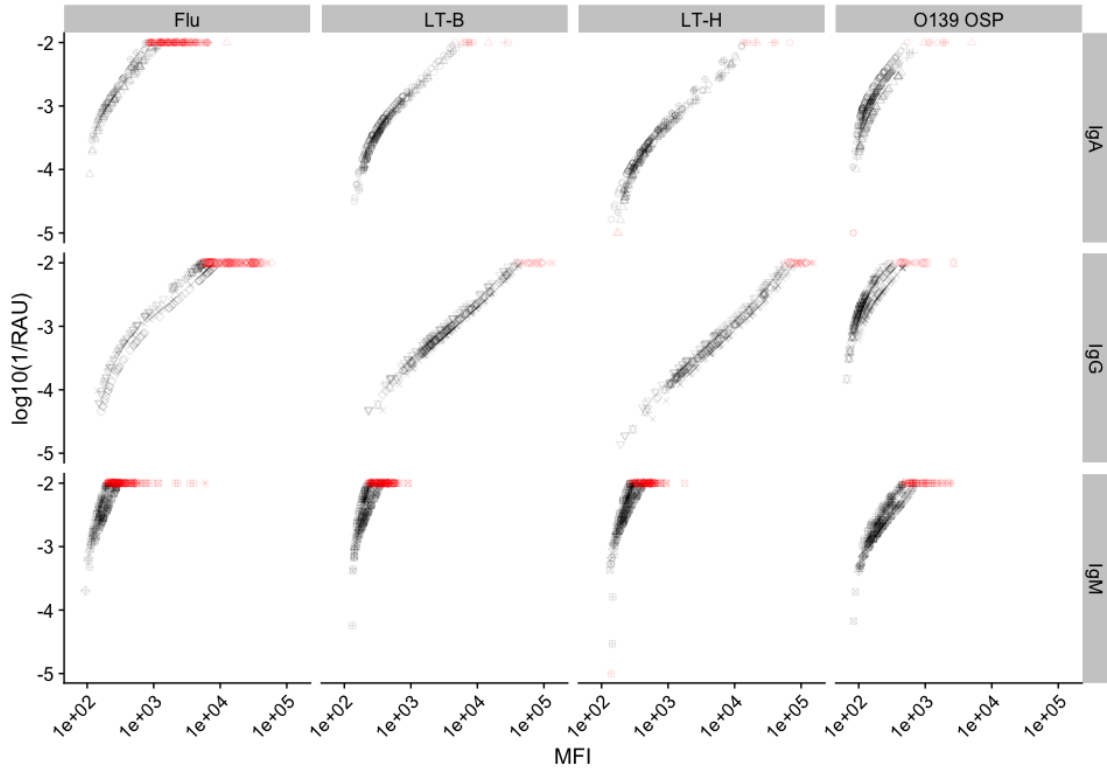
**Figure B.5. Relationship between relative antibody units and median fluorescence intensity for additional antigen multiplex bead assay markers**

Relative antibody unit (RAU) measurements for each sample are plotted against the median fluorescence intensity (MFI) calculated from averaging triplicate measurements. Relative antibody units estimates were truncated at  $10^2$  and  $10^5$  (red points).



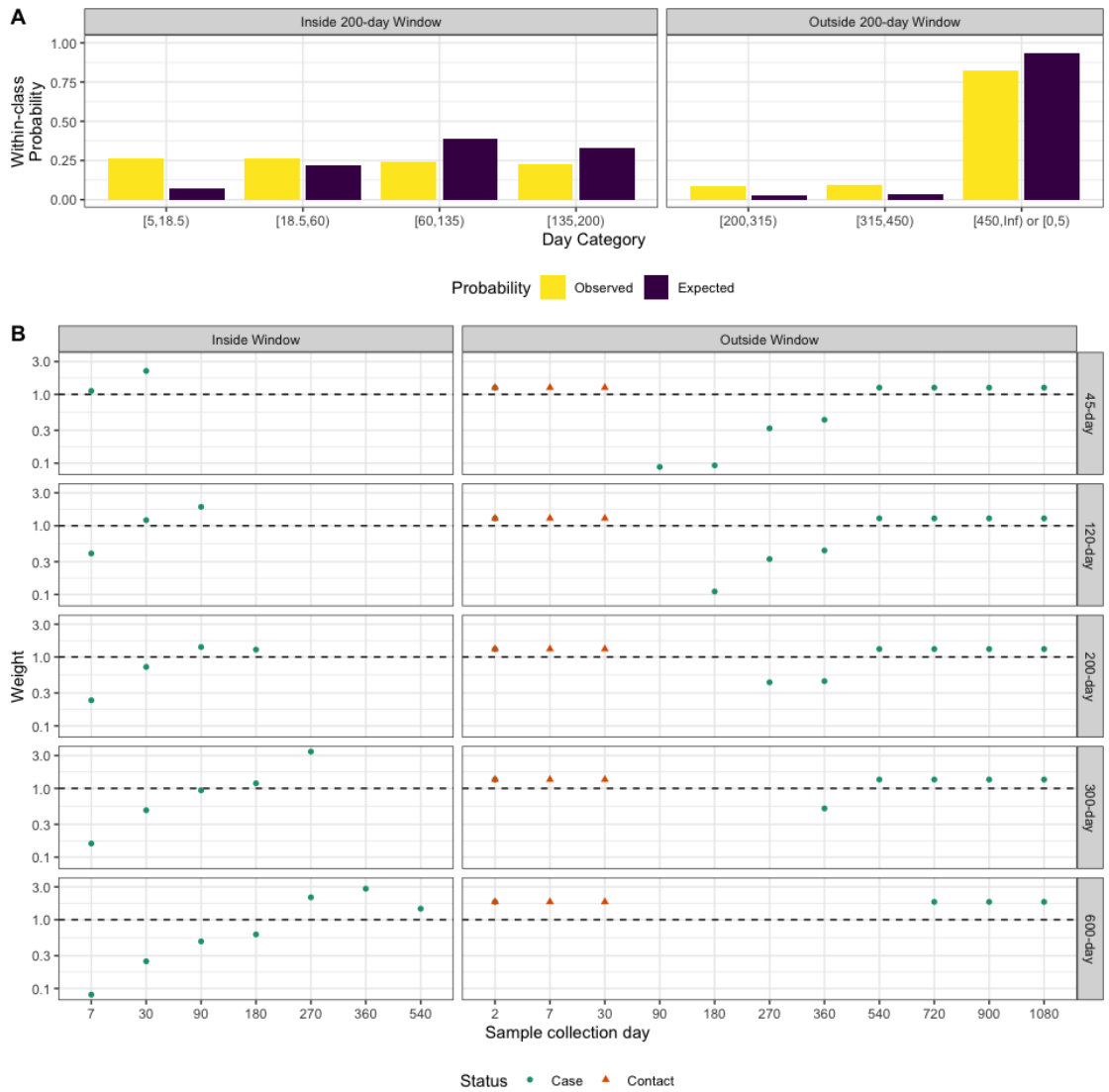
**Figure B.6. Relationship between relative antibody units and median fluorescence intensity for non-*Vibrio cholerae* O1 antigen multiplex bead assay markers**

Relative antibody unit (RAU) measurements for each sample are plotted against the median fluorescence intensity (MFI) calculated from averaging triplicate measurements. Relative antibody units estimates were truncated at  $10^2$  and  $10^5$  (red points). LT-B = heat labile toxin B subunit. LT-H = heat labile toxin holotoxin.



### Figure B.7. Weights used for random forest models at each infection window

The yellow columns represent the proportion of samples that fall within the time interval on the x-axis (A). The purple columns indicate the expected proportion of samples that would fall within the time interval if infection times were exponentially distributed with a 10% annual incidence. Both of these proportions condition on the sample already being either inside or outside the infection window. Each point shows the weight used for each infection window (B).



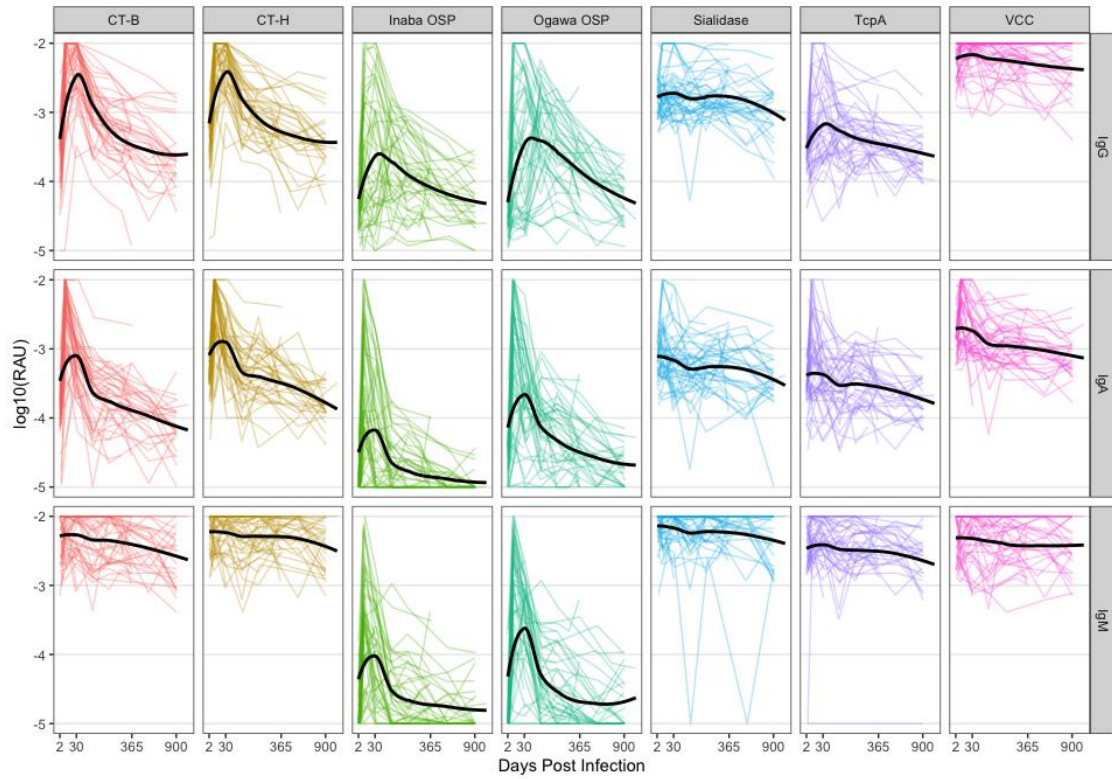
**Table B.1. Individual characteristics of culture confirmed cholera patients and uninfected household contacts**

All cases were hospitalized and solely had O1 *Vibrio cholerae* isolated. Inaba O1 was isolated from non-Ogawa cases

Characteristics	Cholera cases (n=48)	Household contacts (n=3)
Age Group		
< 5 years (%)	8 (17)	0 (0)
5-9 years (%)	14 (29)	0 (0)
10-17 years (%)	9 (19)	2 (67)
18+ years (%)	17 (35)	1 (33)
Female (%)	18 (38)	1 (33)
V. cholerae O1 Ogawa isolated (%)	39 (81)	-

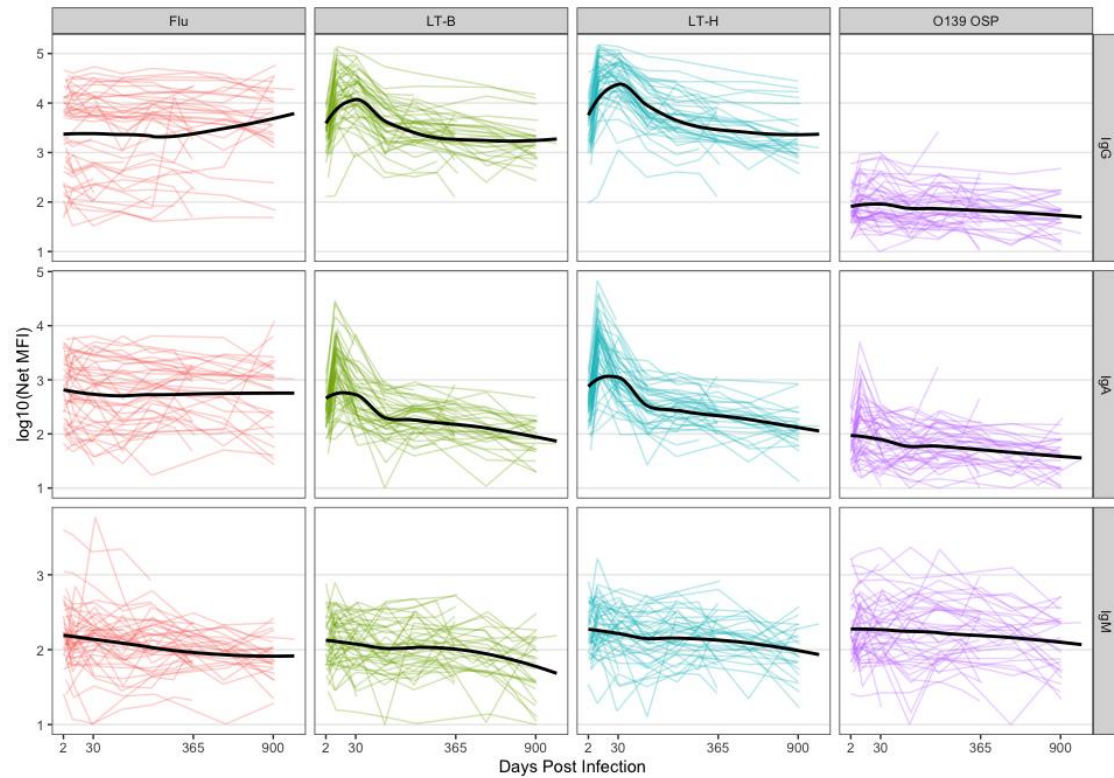
**Figure B.8. Multiplex bead assay relative antibody unit measurements of IgG, IgM, and IgA against *V. cholerae* O1 antigens among culture confirmed cholera patients**

The y-axis indicates the log (base 10) of the relative antibody unit (RAU) and the x-axis is the number of days post-infection (square-root transformed). Each colored line indicates individual trajectories over time. The Black solid line is a loess smooth function.



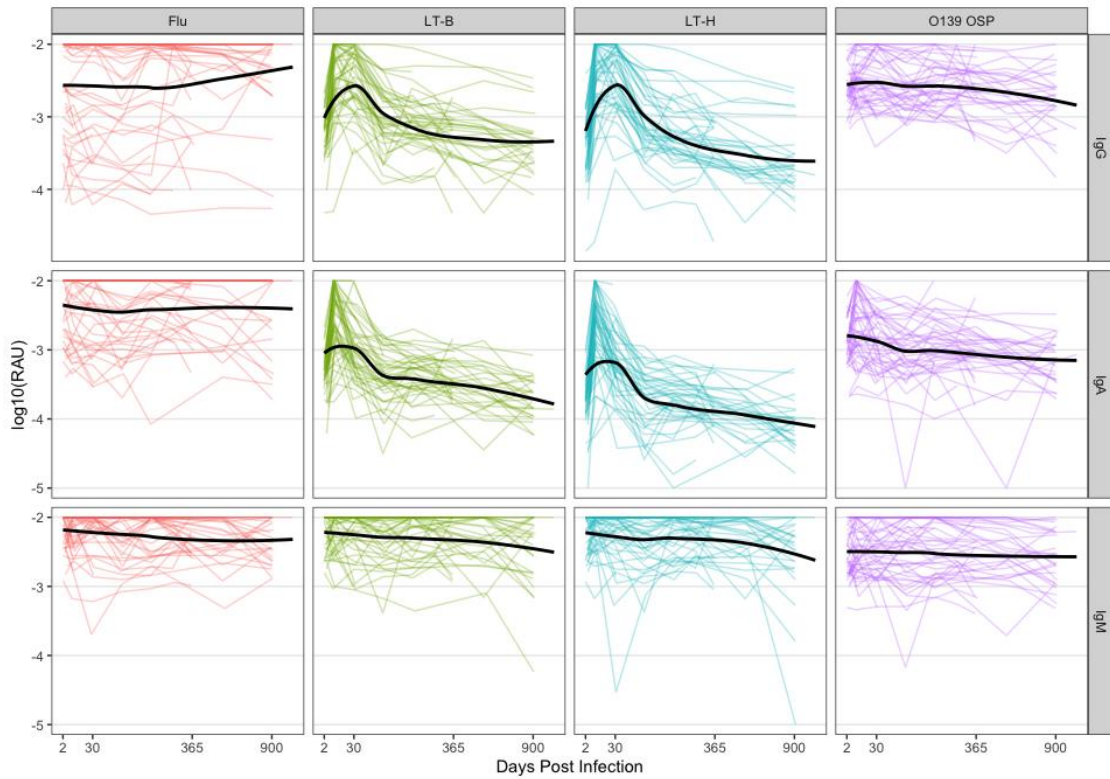
**Figure B.9. Multiplex bead assay Net MFI measurements of IgG, IgM, and IgA against additional antigens among culture confirmed cholera patients**

The y-axis indicates log (base 10) of the Net MFI and the x-axis is the number of days post-infection (square-root transformed). Each colored line indicates individual trajectories over time. The Black solid line is a loess smooth function.

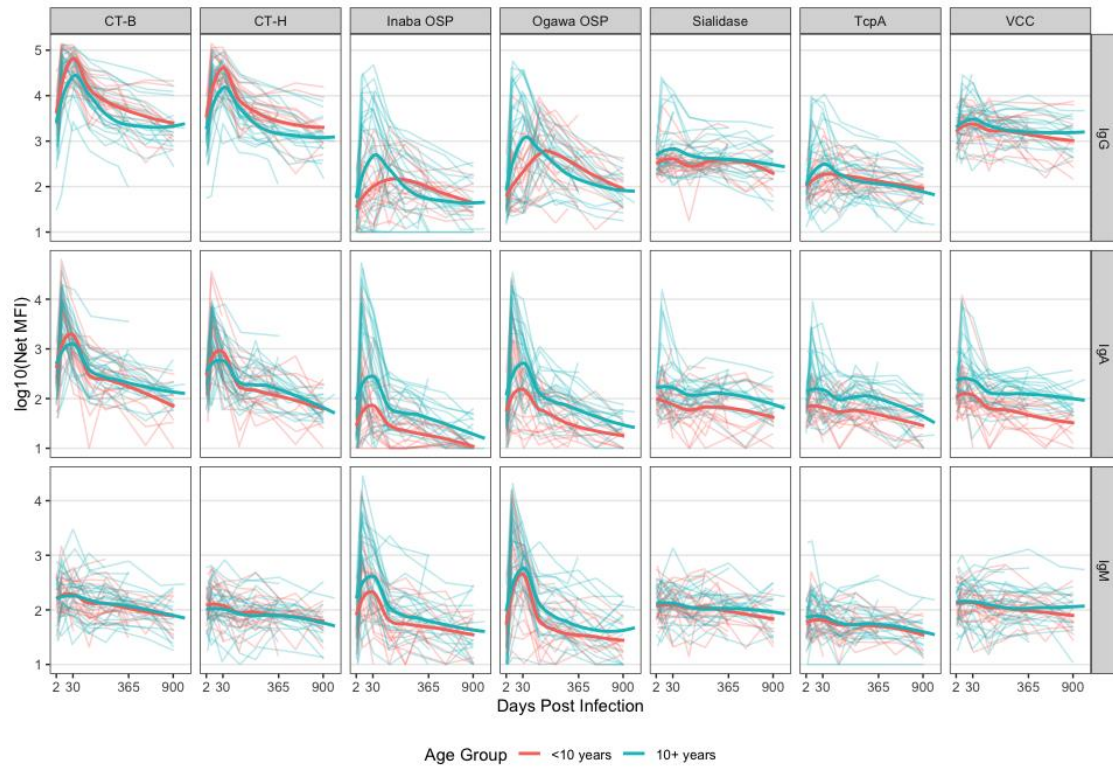


**Figure B.10. Multiplex bead assay relative antibody unit measurements of IgG, IgM, and IgA against additional antigens among culture confirmed cholera patients**

The y-axis indicates the log (base 10) of the relative antibody unit (RAU) and the x-axis is the number of days post-infection (square-root transformed). Each colored line indicates individual trajectories over time. The Black solid line is a loess smooth function.

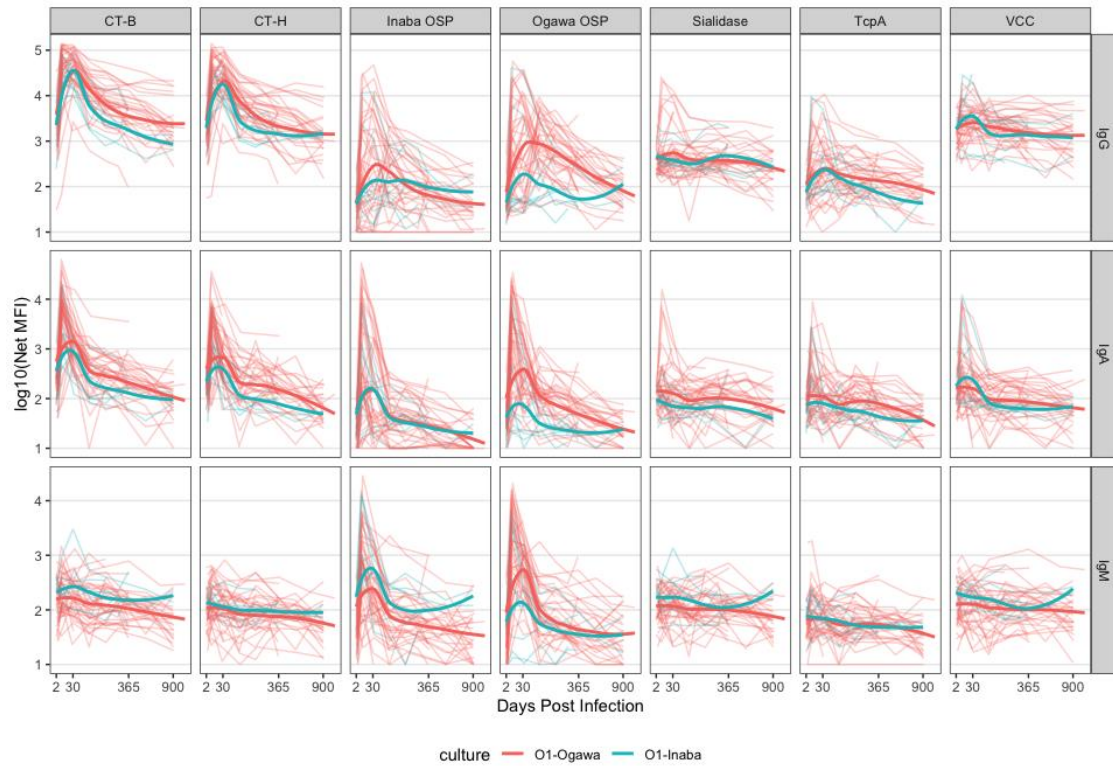


**Figure B.11. Multiplex bead assay Net MFI measurements of IgG, IgM, and IgA against *V. cholerae* O1 antigens among culture confirmed cholera patients, by age group**



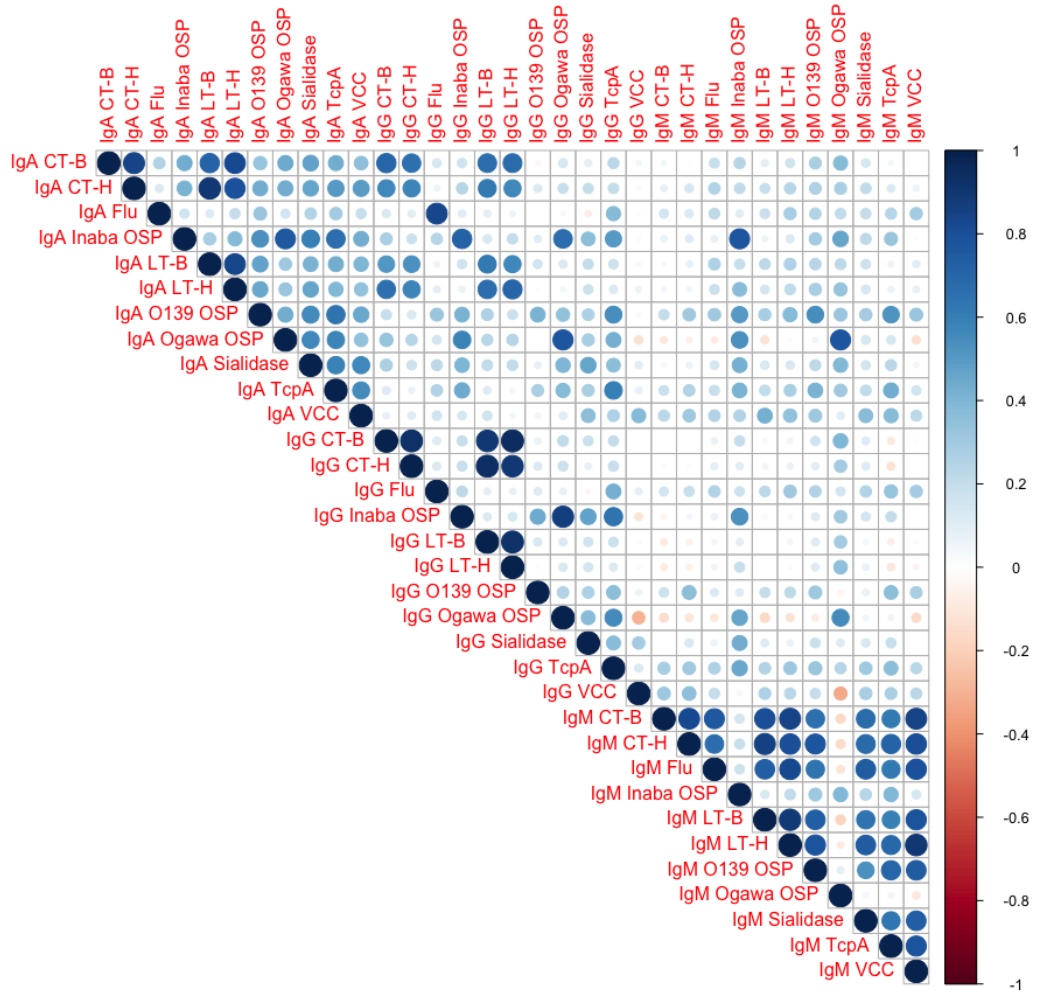


**Figure B.12. Multiplex bead assay Net MFI measurements of IgG, IgM, and IgA against *V. cholerae* O1 antigens among culture confirmed cholera patients, by infecting serotype**



### Figure B.13. Correlation matrix of multiplex bead assay biomarkers

Spearman correlation coefficients were calculated using 49 samples all of which were collected approximately 30 days post-enrollment for both cases and contacts.



**Table B.2. Biphasic and Exponential decay model comparison**

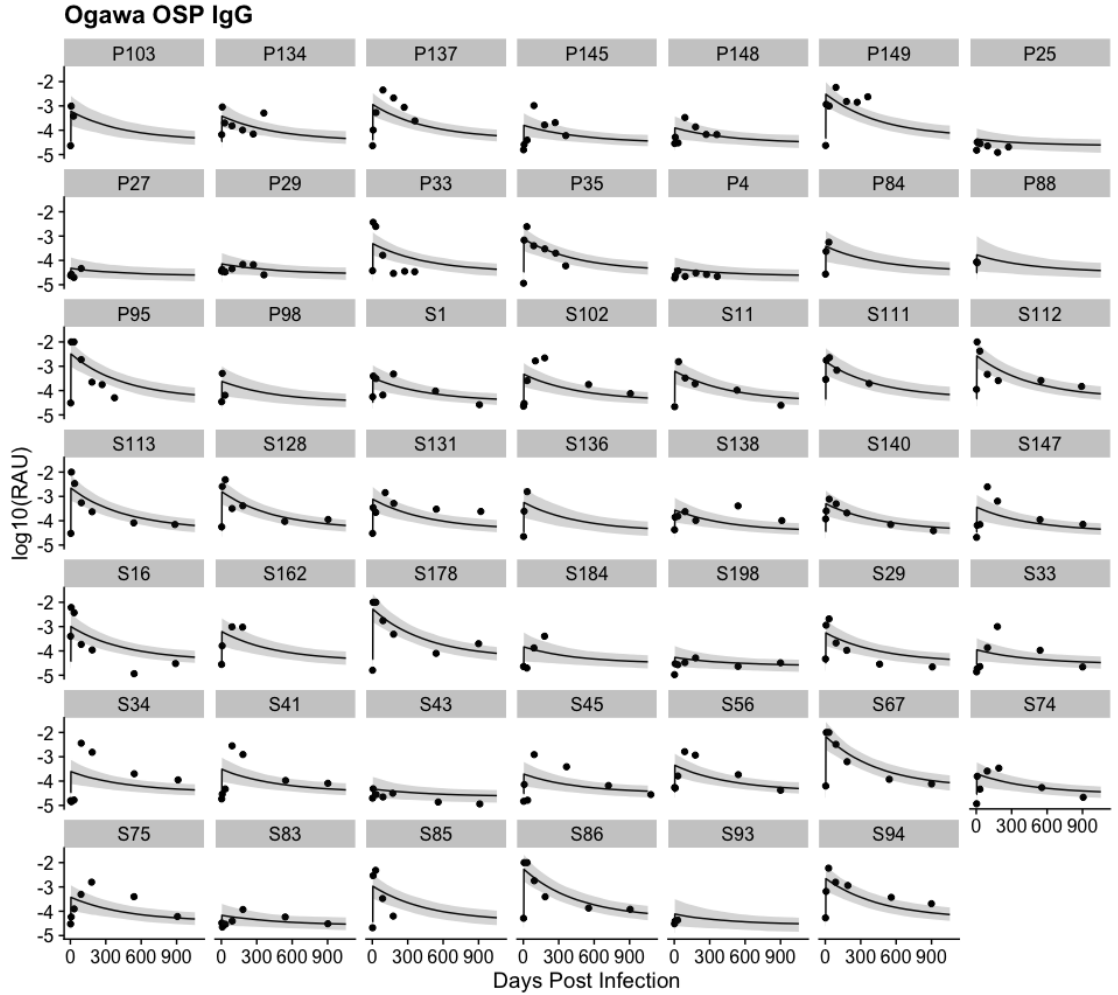
The expected log-predictive density (ELPD) and standard error difference was calculated using the loo R package. Asterisks denote that the ELPD of the biphasic model (which includes more parameters) was over 1.96 standard errors larger than the ELPD of the exponential model.

Marker	Biphasic ELPD	Exponential ELPD	Biphasic - Exponential Standard Error Difference
ELISA CtxB IgA	-94.8	-94.6	-0.62
ELISA CtxB IgG	11.1	11.2	-1.22
ELISA LPS IgA	-135.9	-133.8	-1.37
ELISA LPS IgG	-7.2	-5.2	-0.76
MBA IgA CTHT	-556.8	-394.6	-51.68
MBA IgA CtxB	-266.8	-210.5	-26.73
MBA IgA Flu	-6,497.5	-6,168.7	-99.70
MBA IgA InabaOSPBSA	-1,280.4	-1,283.6	24.69*
MBA IgA LTB	-301.4	-277.8	-10.98
MBA IgA LTh	-153.9	-152.6	-4.78
MBA IgA O139BSA	-85.4	-76.1	-4.86
MBA IgA OgawaOSPBSA	-181.0	-170.7	-6.71
MBA IgA Sialidase	-67.2	-64.3	-1.36
MBA IgA TcpA	-121.9	-109.4	-10.47
MBA IgA VCC	-137.9	-126.8	-8.23
MBA IgG CTHT	-167.7	-159.4	-4.60
MBA IgG CtxB	-153.1	-143.3	-3.57
MBA IgG Flu	-6,379.6	-6,178.0	-95.02
MBA IgG InabaOSPBSA	-186.8	-185.7	-0.80
MBA IgG LTB	-148.9	-159.2	7.61*
MBA IgG LTh	-256.5	-247.0	-9.61
MBA IgG O139BSA	-13.0	-4.9	-5.52
MBA IgG OgawaOSPBSA	-248.2	-247.5	-0.75
MBA IgG Sialidase	-7.5	-5.9	-0.78
MBA IgG TcpA	-107.5	-107.3	-0.87

Marker	Biphasic ELPD	Exponential ELPD	Biphasic - Exponential Standard Error Difference
MBA IgG VCC	-393.2	-437.8	16.36*
MBA IgM CTHT	-274.4	-261.3	-13.15
MBA IgM CtxB	-460.6	-465.8	20.42*
MBA IgM Flu	-444.3	-498.3	20.54*
MBA IgM InabaOSPBSA	-780.1	-798.8	11.13*
MBA IgM LTB	-628.5	-595.8	-22.63
MBA IgM LTh	-226.7	-243.4	11.18*
MBA IgM O139BSA	-60.8	-56.8	-8.01
MBA IgM OgawaOSPBSA	-367.4	-337.7	-11.41
MBA IgM Sialidase	-337.1	-339.8	5.66*
MBA IgM TcpA	-606.0	-630.2	19.58*
MBA IgM VCC	-933.2	-964.8	26.80*
Vibriocidal Inaba	-616.5	-587.1	-14.17
Vibriocidal Ogawa	-675.2	-617.3	-28.75

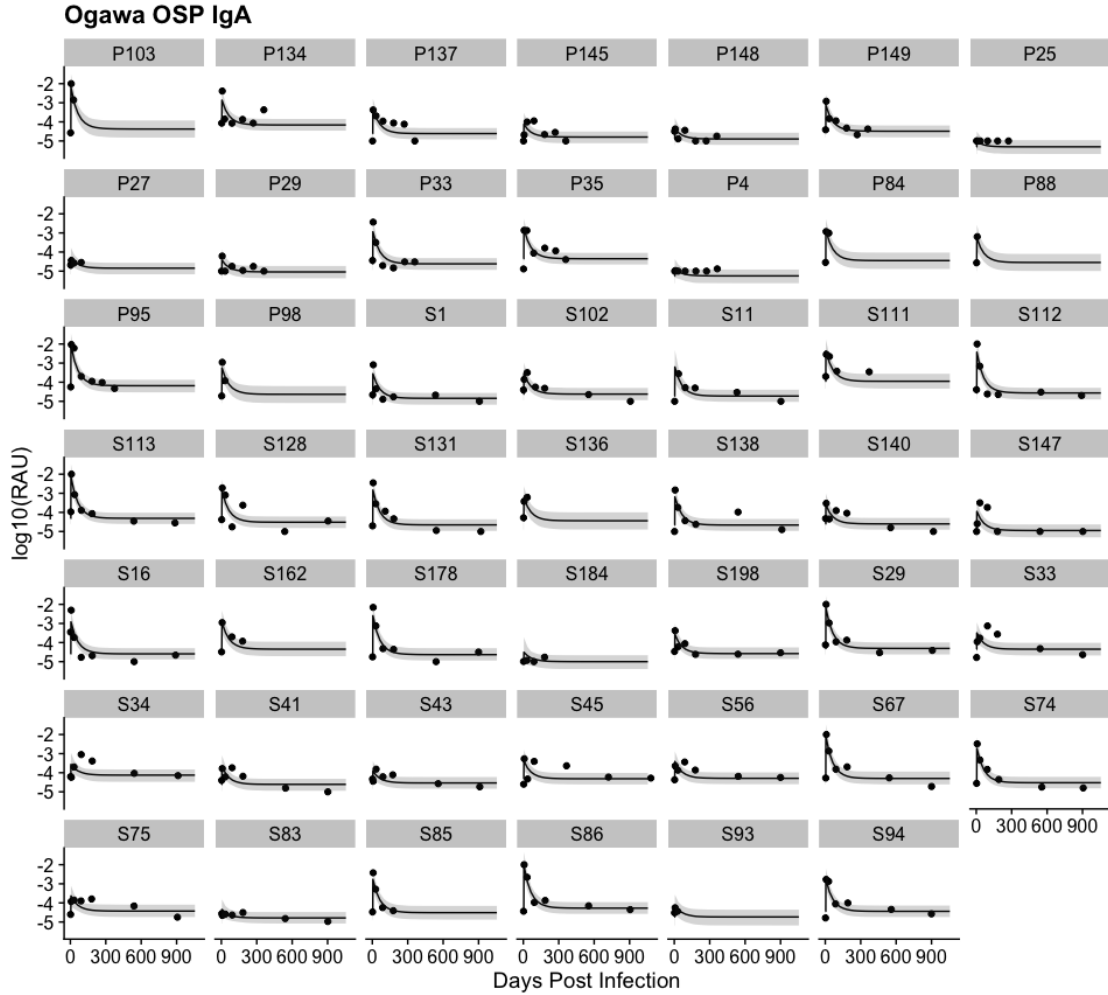
### Figure B.14. Individual-level trajectories of Ogawa OSP IgG

Each facet shows the  $\log_{10}(\text{RAU})$  measurements for individuals over time (points). Solid line indicates the median value of exponential decay model. Shaded area is the 95% credible interval.



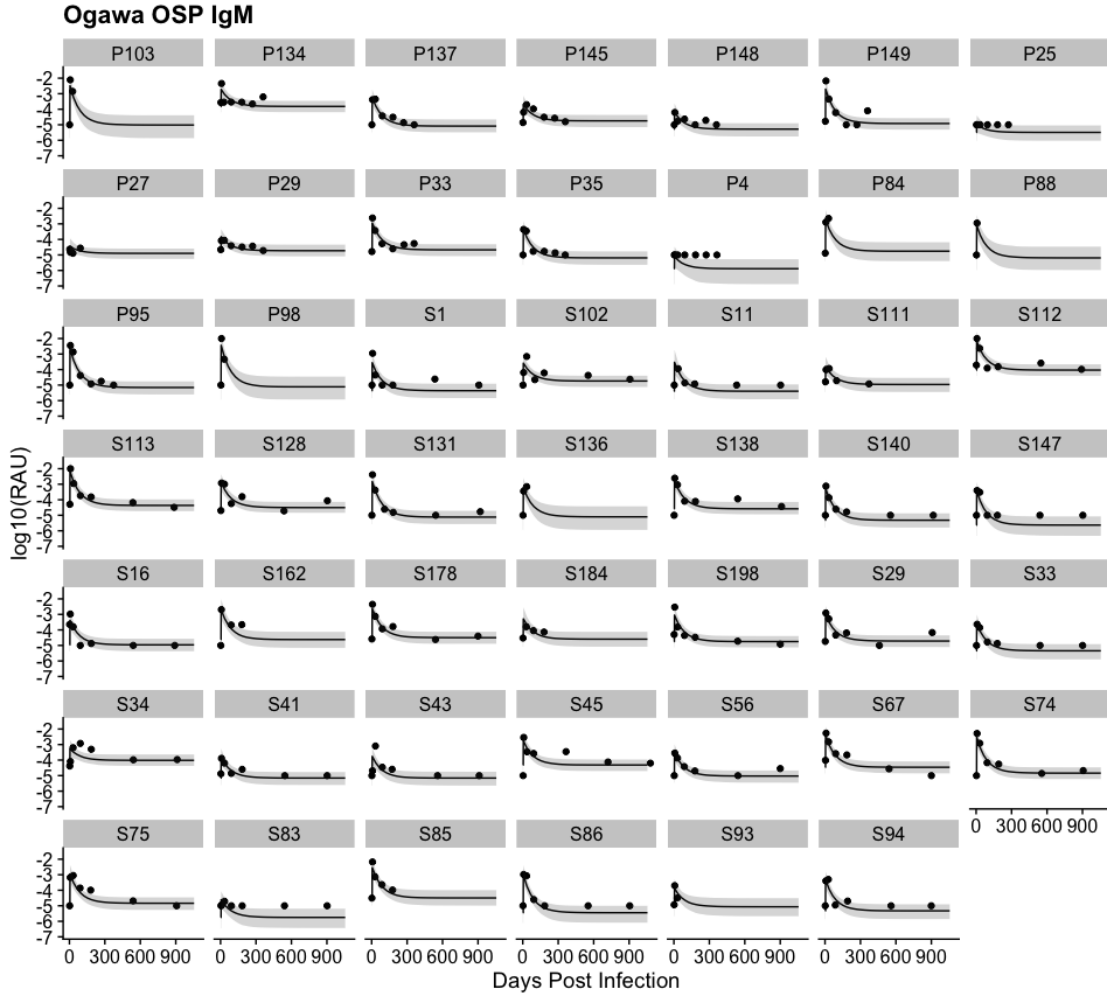
### Figure B.15. Individual-level trajectories of Ogawa OSP IgA

Each facet shows the  $\log_{10}(\text{RAU})$  measurements for individuals over time (points). Solid line indicates the median value of exponential decay model. Shaded area is the 95% credible interval.



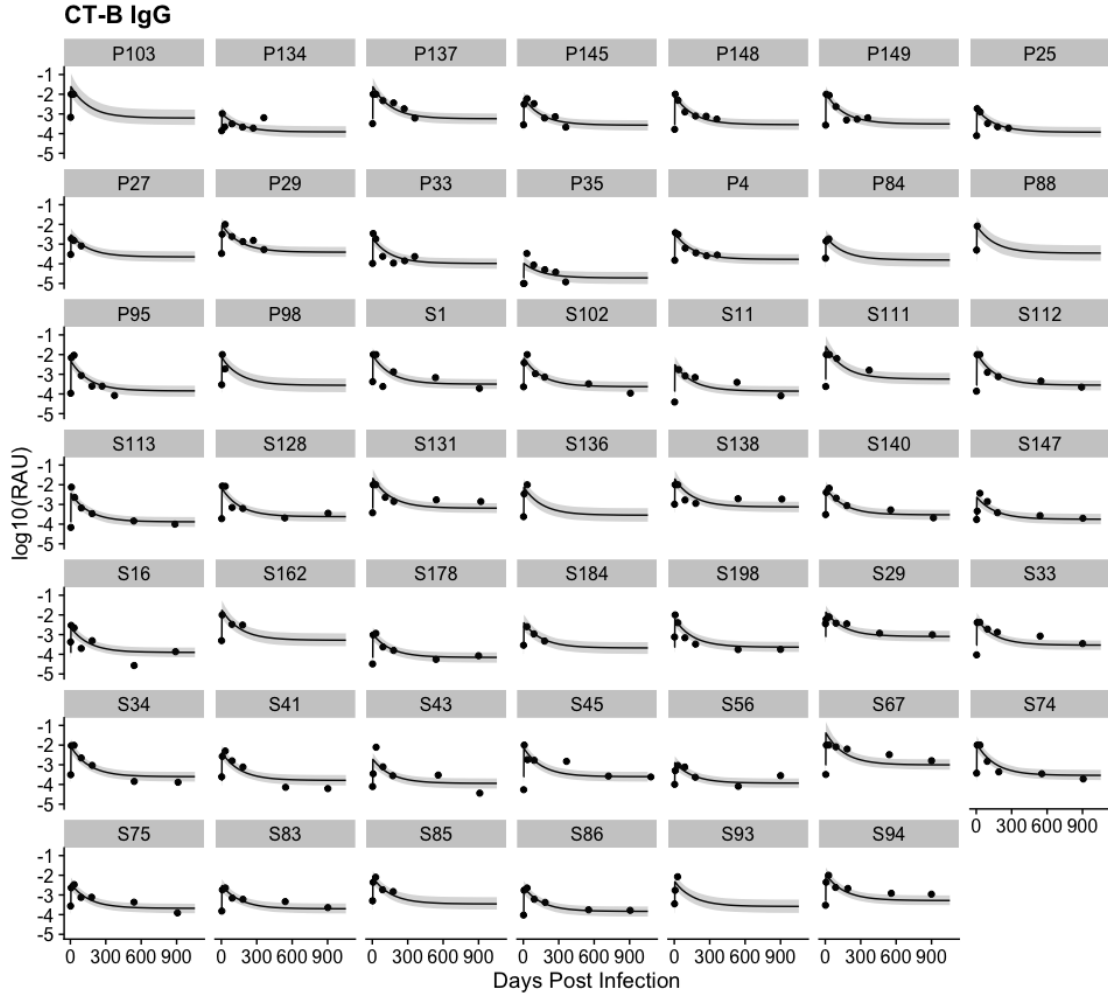
### Figure B.16. Individual-level trajectories of Ogawa OSP IgM

Each facet shows the  $\log_{10}(\text{RAU})$  measurements for individuals over time (points). Solid line indicates the median value of exponential decay model. Shaded area is the 95% credible interval.



### Figure B.17. Individual-level trajectories of CT-B IgG

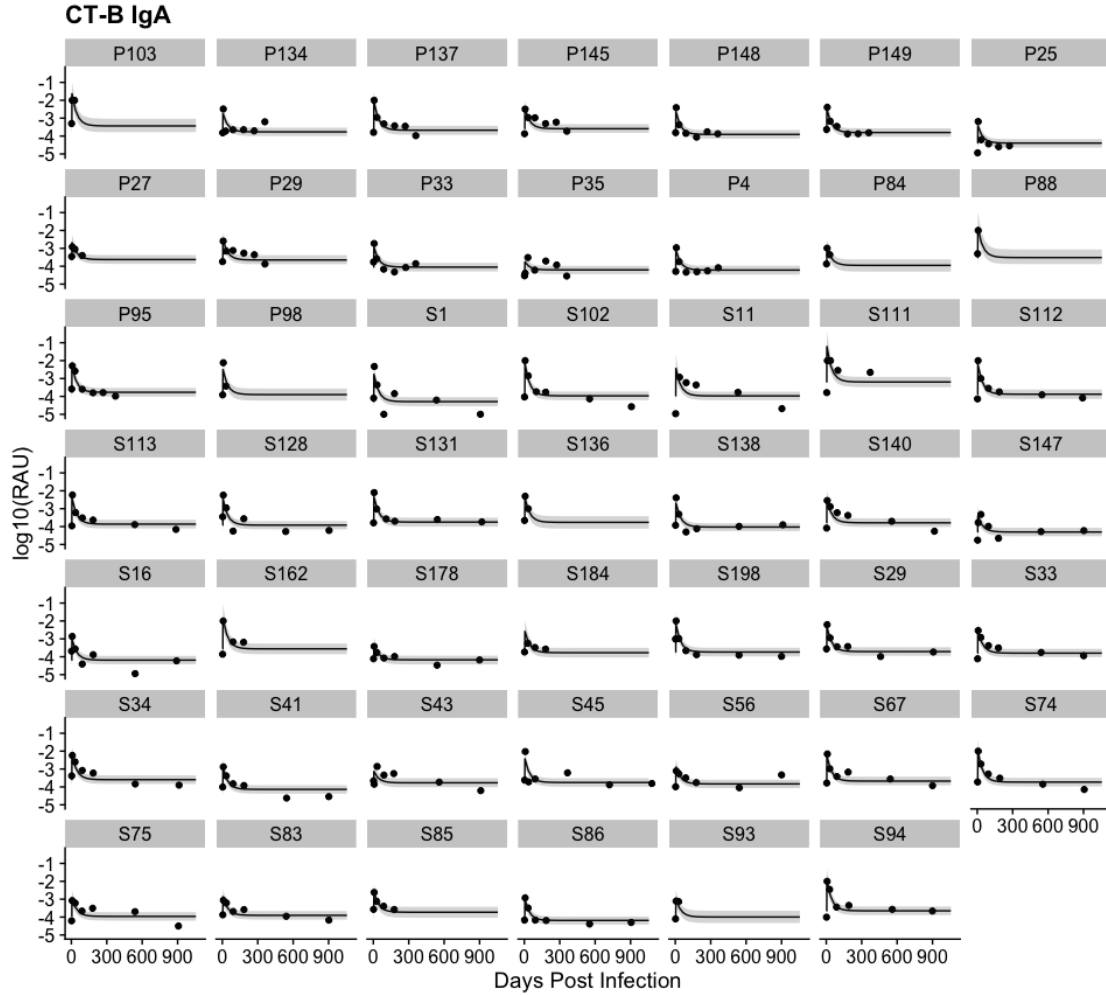
Each facet shows the  $\log_{10}(\text{RAU})$  measurements for individuals over time (points). Solid line indicates the median value of exponential decay model. Shaded area is the 95% credible interval.





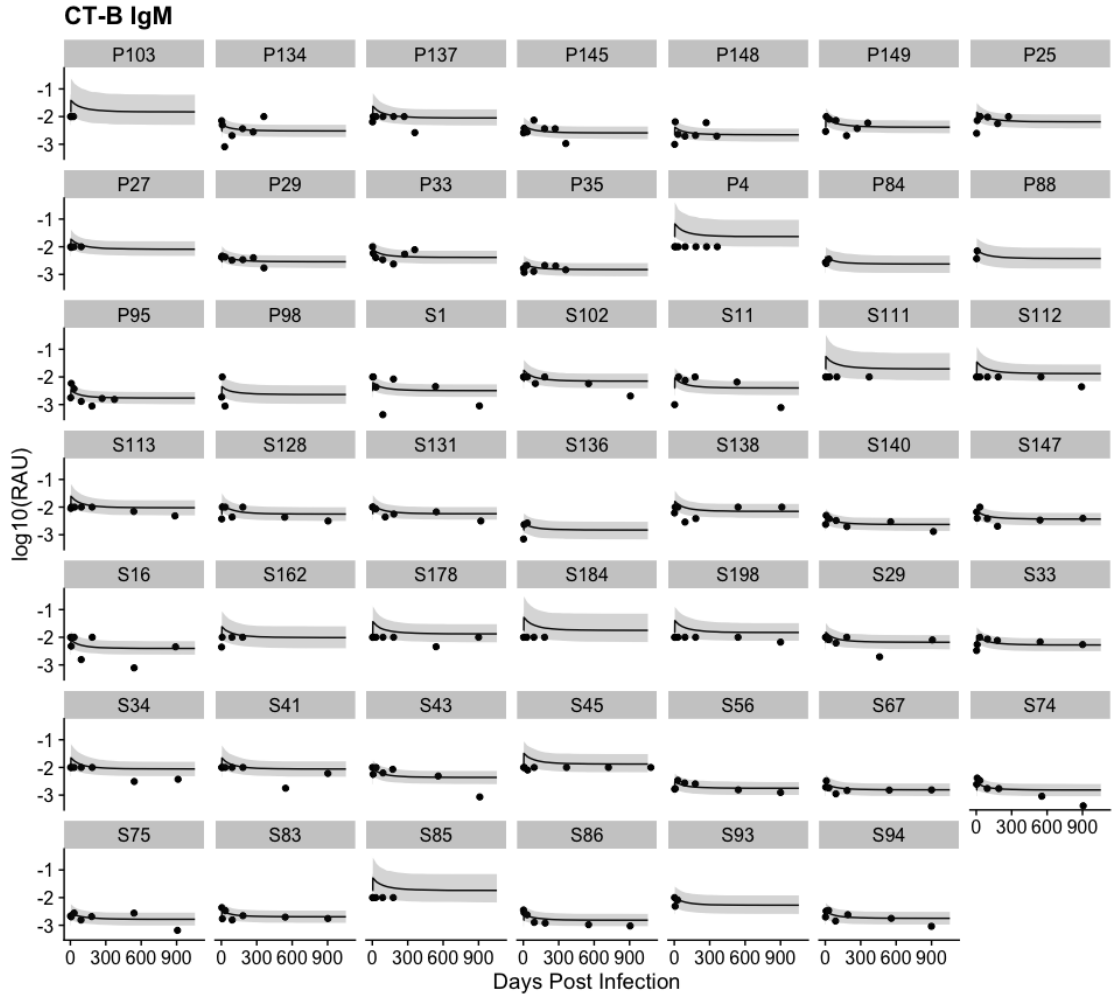
### Figure B.18. Individual-level trajectories of CT-B IgA

Each facet shows the  $\log_{10}(\text{RAU})$  measurements for individuals over time (points). Solid line indicates the median value of exponential decay model. Shaded area is the 95% credible interval.



### Figure B.19. Individual-level trajectories of CT-B IgM

Each facet shows the  $\log_{10}(\text{RAU})$  measurements for individuals over time (points). Solid line indicates the median value of exponential decay model. Shaded area is the 95% credible interval.



**Table B.3. Estimated duration of half-life and average fold-change from univariate exponential decay models**

The median estimate duration of half-life (in days) and the median estimate increase in fold-change for the average individual are shown below as well as in Figure 2. 95% Bayesian Credible Intervals are shown in parentheses.

Isotype	Antigen	Half Life (95% CI) in days	Average Fold-Change (95% CI)
IgA	CT-H	21 (16-27)	21.2 (13.1-15.8)
	CT-B	29 (22-38)	28.8 (18.9-22.3)
	Flu	11 (5-39)	2.0 (1.3-4.8)
	Inaba OSP	33 (26-43)	32.7 (16.1-26.4)
	LT-B	20 (16-25)	13.0 (8.6-15.6)
	LT-H	22 (18-28)	21.7 (14.8-18.2)
	O139 OSP	12 (6-24)	2.8 (2.1-6.2)
	Ogawa OSP	42 (31-59)	25.3 (14.3-30.8)
	Sialidase	16 (7-37)	1.9 (1.5-6.8)
	TcpA	18 (11-30)	3.5 (2.5-11.2)
	VCC	19 (13-29)	3.3 (2.5-13.2)
IgG	CT-H	98 (77-125)	16.0 (11.7-77.4)
	CT-B	108 (86-135)	24.6 (18.1-86.2)
	Flu	11 (5-71)	1.6 (1.2-5.0)
	Inaba OSP	130 (86-215)	6.7 (4.2-85.8)
	LT-B	66 (51-85)	8.4 (6.2-51.4)
	LT-H	77 (63-94)	15.2 (10.8-62.6)
	O139 OSP	86 (19-234)	1.6 (1.3-19.2)
	Ogawa OSP	335 (221-490)	12.7 (7.6-221.3)
	Sialidase	75 (19-260)	1.5 (1.3-19.4)
	TcpA	133 (77-242)	3.1 (2.3-77.4)
	VCC	78 (32-162)	1.9 (1.5-31.5)
IgM	CT-H	23 (9-92)	2.1 (1.6-8.8)
	CT-B	68 (17-178)	2.3 (1.7-17.2)
	Flu	53 (9-192)	1.8 (1.4-9.2)
	Inaba OSP	36 (28-46)	24.0 (13.4-28.0)

Isotype	Antigen	Half Life (95% CI) in days	Average Fold-Change (95% CI)
	LT-B	18 (6-100)	1.7 (1.3-6.2)
	LT-H	15 (5-173)	2.0 (1.4-5.4)
	O139 OSP	61 (5-386)	1.4 (1.2-5.5)
	Ogawa OSP	59 (46-76)	62.1 (34.6-46.2)
	Sialidase	25 (7-130)	1.9 (1.4-6.9)
	TcpA	69 (13-363)	1.8 (1.4-12.7)
	VCC	30 (9-90)	1.7 (1.3-8.9)
Marker		Half Life (95% CI) in days	Average Fold-Change (95% CI)
	ELISA CtxB IgA	17 (12-22)	2.0 (1.8-12.3)
	ELISA CtxB IgG	76 (53-112)	1.5 (1.4-53.1)
	ELISA LPS IgA	36 (25-53)	1.9 (1.6-25.5)
	ELISA LPS IgG	56 (36-87)	1.4 (1.3-36.2)
	Vibriocidal Inaba	72 (49-107)	49.7 (27.8-49.2)
	Vibriocidal Ogawa	142 (85-251)	52.0 (26.1-85.4)

**Table B.4. Relative parameter values for univariate exponential decay models including covariates for different serological markers**

The median relative baseline value, median relative fold-change value, and median difference in days of half-life are reported. 95% Bayesian Credible intervals are shown in parentheses. Reference categories include male, non-O blood group, Inaba, and 10+ years old.

Marker	Covariate	Ratio of Baseline Values	Ratio of Fold-Change Values	Difference in days of Half-Life
ELISA IgA CtxB	Female	0.73 (0.488-1.09)	0.996 (0.484-1.97)	-3 (-12-6)
	O Blood	0.875 (0.587-1.32)	1.77 (0.923-3.26)	8 (-1-17)
	Ogawa	1.16 (0.68-1.99)	0.848 (0.348-1.81)	3 (-9-12)
	Under 10	0.817 (0.547-1.22)	2.17 (1.13-4.05)*	0 (-10-9)
ELISA IgA LPS	Female	1.16 (0.774-1.74)	0.333 (0.14-0.708)*	-6 (-26-32)
	O Blood	1.35 (0.908-2)	1.23 (0.56-2.46)	-12 (-55-16)
	Ogawa	1.36 (0.851-2.24)	0.509 (0.188-1.14)	19 (-4-40)
	Under 10	0.477 (0.337-0.669)*	0.552 (0.251-1.12)	21 (-9-70)
ELISA IgG CtxB	Female	0.993 (0.742-1.33)	1.18 (0.83-1.66)	-42 (-114-8)
	O Blood	1.04 (0.782-1.38)	1.13 (0.805-1.6)	-22 (-95-30)
	Ogawa	0.913 (0.64-1.32)	0.754 (0.476-1.19)	49 (4-106)*
	Under 10	1.2 (0.897-1.6)	1.28 (0.904-1.8)	-27 (-107-24)
ELISA IgG LPS	Female	0.938 (0.717-1.23)	0.765 (0.482-1.15)	16 (-30-98)
	O Blood	0.998 (0.767-1.33)	1.21 (0.792-1.78)	-28 (-201-31)
	Ogawa	1.05 (0.75-1.51)	0.738 (0.418-1.21)	9 (-64-53)
	Under 10	0.829 (0.638-1.07)	0.61 (0.385-0.91)*	46 (-16-200)
MBA IgA CTHT	Female	1.11 (0.713-1.72)	0.536 (0.191-1.33)	2 (-9-20)
	O Blood	1.12 (0.74-1.69)	0.747 (0.275-1.8)	8 (-2-22)
	Ogawa	1.4 (0.847-2.32)	1.02 (0.33-2.39)	10 (-1-19)
	Under 10	0.921 (0.607-1.39)	1.42 (0.526-3.41)	-2 (-13-8)
MBA IgA CtxB	Female	1.17 (0.721-1.91)	0.495 (0.206-1.16)	15 (-3-44)
	O Blood	1.09 (0.681-1.75)	0.975 (0.41-2.24)	3 (-12-18)
	Ogawa	1.26 (0.701-2.33)	1.49 (0.528-3.65)	14 (-2-27)
	Under 10	0.903 (0.56-1.46)	1.41 (0.601-3.24)	-6 (-21-9)
	Female	2.57 (0.76-9.56)	0.82 (0.396-1.86)	-2 (-22-57)

Marker	Covariate	Ratio of Baseline Values	Ratio of Fold-Change Values	Difference in days of Half-Life
MBA IgA Flu	O Blood	1.09 (0.316-3.41)	0.716 (0.327-1.57)	-3 (-51-21)
	Ogawa	1.61 (0.341-9.15)	0.683 (0.252-1.41)	-2 (-31-37)
	Under 10	0.32 (0.0826-0.922)*	0.781 (0.272-1.57)	-3 (-180-25)
MBA IgA Inaba OSP	Female	2.59 (1.13-6.46)*	0.404 (0.102-1.31)	-15 (-31-1)
	O Blood	1.48 (0.645-3.54)	1.98 (0.611-5.5)	10 (-7-26)
	Ogawa	1.47 (0.469-5)	0.333 (0.0629-1.14)	10 (-12-25)
	Under 10	0.325 (0.147-0.663)*	0.156 (0.0352-0.584)*	-2 (-19-23)
MBA IgA LTB	Female	0.796 (0.534-1.18)	0.666 (0.274-1.49)	3 (-7-20)
	O Blood	1.29 (0.888-1.89)	0.801 (0.356-1.67)	5 (-5-16)
	Ogawa	1.26 (0.776-2.06)	1.14 (0.44-2.39)	2 (-13-12)
	Under 10	0.933 (0.632-1.38)	1.4 (0.627-2.86)	0 (-9-10)
MBA IgA LTh	Female	1.07 (0.665-1.7)	0.592 (0.272-1.28)	6 (-5-24)
	O Blood	1.37 (0.878-2.16)	0.743 (0.341-1.53)	5 (-4-16)
	Ogawa	1.43 (0.775-2.49)	1.21 (0.469-2.76)	4 (-9-13)
	Under 10	0.851 (0.545-1.34)	1.57 (0.728-3.32)	1 (-9-10)
MBA IgA O139	Female	1.11 (0.703-1.76)	0.432 (0.227-0.792)*	-4 (-13-37)
	O Blood	1.19 (0.777-1.86)	0.616 (0.324-1.18)	-2 (-13-19)
	Ogawa	1.18 (0.627-2.14)	1.15 (0.551-2.18)	-4 (-56-8)
	Under 10	0.485 (0.334-0.702)*	0.831 (0.418-1.56)	-4 (-24-7)
MBA IgA Ogawa OSP	Female	1.48 (0.812-2.68)	0.334 (0.0942-1.01)	-9 (-33-23)
	O Blood	1.31 (0.726-2.36)	1.48 (0.517-3.68)	-11 (-43-16)
	Ogawa	3 (1.58-5.81)*	1.17 (0.342-3.05)	26 (1-48)*
	Under 10	0.377 (0.216-0.652)*	0.273 (0.0803-0.809)*	33 (-4-83)
MBA IgA Sialidase	Female	1.01 (0.657-1.54)	0.72 (0.429-1.24)	-5 (-23-32)
	O Blood	1.22 (0.813-1.86)	0.862 (0.501-1.43)	4 (-15-42)
	Ogawa	1.47 (0.866-2.42)	0.812 (0.411-1.44)	3 (-22-28)
	Under 10	0.651 (0.451-0.961)*	0.752 (0.443-1.31)	-9 (-38-9)
MBA IgA TcpA	Female	1.11 (0.72-1.73)	0.554 (0.26-1.1)	-5 (-19-15)
	O Blood	1.07 (0.686-1.67)	0.537 (0.24-1.06)	15 (2-36)*
	Ogawa	1.44 (0.846-2.5)	0.267 (0.097-0.555)*	9 (-2-23)

Marker	Covariate	Ratio of Baseline Values	Ratio of Fold-Change Values	Difference in days of Half-Life
	Under 10	0.5 (0.352-0.709)*	0.336 (0.158-0.665)*	-9 (-21-7)
MBA IgA VCC	Female	1.21 (0.81-1.83)	0.941 (0.491-1.62)	10 (-4-31)
	O Blood	1.04 (0.699-1.54)	0.646 (0.338-1.09)	9 (-4-26)
	Ogawa	1.09 (0.647-1.8)	0.428 (0.175-0.798)*	3 (-12-14)
	Under 10	0.639 (0.443-0.924)*	0.568 (0.289-0.998)*	-6 (-19-8)
MBA IgG CTHT	Female	0.541 (0.332-0.878)*	1.48 (0.769-2.78)	28 (-17-77)
	O Blood	0.965 (0.583-1.62)	0.794 (0.426-1.47)	-7 (-53-42)
	Ogawa	0.883 (0.479-1.63)	1.75 (0.835-3.48)	59 (12-98)*
	Under 10	1.92 (1.2-3.05)*	1.16 (0.621-2.21)	-34 (-87-11)
MBA IgG CtxB	Female	0.807 (0.444-1.42)	1.07 (0.553-2.03)	32 (-13-84)
	O Blood	0.686 (0.389-1.21)	1.05 (0.563-1.93)	10 (-35-60)
	Ogawa	1.26 (0.618-2.61)	1.25 (0.593-2.61)	44 (-8-88)
	Under 10	1.66 (0.93-2.93)	1.05 (0.559-1.99)	-20 (-70-25)
MBA IgG Flu	Female	1.41 (0.306-6.22)	0.836 (0.483-1.57)	-2 (-67-55)
	O Blood	1.3 (0.292-5.02)	0.671 (0.38-1.18)	-4 (-111-81)
	Ogawa	4.46 (0.76-27.6)	0.64 (0.299-1.14)	-2 (-33-87)
	Under 10	0.347 (0.0889-1.51)	0.68 (0.352-1.27)	-3 (-109-58)
MBA IgG Inaba OSP	Female	1.31 (0.738-2.32)	0.223 (0.0859-0.567)*	-76 (-162--12)*
	O Blood	1.65 (0.908-3.05)	0.857 (0.328-2.03)	-116 (-336-14)
	Ogawa	0.971 (0.457-2.43)	0.306 (0.102-0.7)*	20 (-355-100)
	Under 10	0.84 (0.451-1.84)	0.198 (0.0634-0.525)*	333 (-77-945)
MBA IgG LTB	Female	0.6 (0.361-0.986)*	1.21 (0.638-2.19)	24 (-8-60)
	O Blood	0.877 (0.532-1.44)	0.938 (0.513-1.61)	1 (-32-35)
	Ogawa	1.32 (0.723-2.47)	1.29 (0.632-2.36)	34 (-2-63)
	Under 10	1.23 (0.756-2.01)	1.53 (0.853-2.68)	-17 (-54-15)
MBA IgG LTh	Female	0.727 (0.398-1.33)	1.06 (0.498-2.15)	28 (-4-65)
	O Blood	0.979 (0.535-1.76)	0.903 (0.455-1.75)	4 (-27-37)
	Ogawa	1.24 (0.575-2.62)	1.46 (0.653-2.86)	34 (-2-62)
	Under 10	1.5 (0.843-2.72)	1.55 (0.793-3.03)	-11 (-44-20)
	Female	0.583 (0.252-0.892)*	0.843 (0.563-1.22)	0 (-125-237)

Marker	Covariate	Ratio of Baseline Values	Ratio of Fold-Change Values	Difference in days of Half-Life
MBA IgG O139	O Blood	1.16 (0.743-1.8)	0.649 (0.453-0.981)*	-105 (-240-37)
	Ogawa	1.08 (0.619-1.89)	0.741 (0.423-1.18)	75 (-68-219)
	Under 10	0.876 (0.566-1.34)	0.795 (0.549-1.15)	-80 (-262-59)
MBA IgG Ogawa OSP	Female	0.971 (0.46-2.08)	0.444 (0.135-1.23)	7 (-255-343)
	O Blood	1.53 (0.753-3.09)	1.45 (0.471-3.81)	-280 (-681--20)*
	Ogawa	1.8 (0.892-4.08)	0.962 (0.284-2.52)	220 (-8-400)
MBA IgG Sialidase	Under 10	0.775 (0.41-1.46)	0.254 (0.08-0.732)*	645 (262-1857)*
	Female	0.902 (0.588-1.39)	0.848 (0.591-1.23)	33 (-78-555)
	O Blood	1.24 (0.826-1.91)	0.765 (0.542-1.09)	-41 (-245-134)
MBA IgG TcpA	Ogawa	0.756 (0.462-1.29)	0.889 (0.551-1.24)	73 (-12-217)
	Under 10	0.819 (0.549-1.22)	0.673 (0.488-0.948)*	-72 (-238-51)
	Female	1.28 (0.805-2.05)	0.553 (0.27-1.03)	-31 (-132-150)
MBA IgG VCC	O Blood	0.993 (0.64-1.53)	0.473 (0.234-0.885)*	-152 (-408--17)*
	Ogawa	1.39 (0.821-2.48)	0.252 (0.102-0.515)*	37 (-45-133)
	Under 10	0.829 (0.521-1.32)	0.334 (0.173-0.623)*	67 (-79-345)
MBA IgM CHTT	Female	0.898 (0.539-1.46)	1.45 (0.95-2.15)	48 (-98-150)
	O Blood	0.817 (0.497-1.32)	0.775 (0.479-1.18)	-69 (-199-29)
	Ogawa	0.924 (0.505-1.69)	0.562 (0.269-0.976)*	66 (-13-173)
MBA IgM CtxB	Under 10	0.749 (0.457-1.19)	0.855 (0.529-1.3)	-20 (-157-103)
	Female	1.42 (0.896-2.29)	0.734 (0.388-1.31)	-5 (-39-99)
	O Blood	0.745 (0.459-1.21)	0.661 (0.337-1.19)	-8 (-59-107)
MBA IgM Flu	Ogawa	0.455 (0.253-0.835)*	0.734 (0.326-1.29)	10 (-32-104)
	Under 10	0.986 (0.603-1.62)	1.2 (0.653-2.15)	-33 (-225-7)
	Female	1.52 (0.796-2.87)	0.63 (0.312-1.16)	11 (-68-188)
MBA IgM Flu	O Blood	0.599 (0.32-1.09)	0.901 (0.458-1.62)	-21 (-180-107)
	Ogawa	0.338 (0.148-0.72)*	1.04 (0.499-1.81)	8 (-283-120)
	Under 10	1.06 (0.551-1.99)	1.08 (0.583-1.91)	-72 (-248-33)
MBA IgM Flu	Female	1.85 (1.01-3.38)*	0.795 (0.413-1.5)	-2 (-95-166)
	O Blood	0.607 (0.341-1.1)	0.692 (0.362-1.32)	-13 (-160-142)
	Ogawa	0.49 (0.238-1.02)	0.47 (0.171-0.955)*	28 (-128-158)

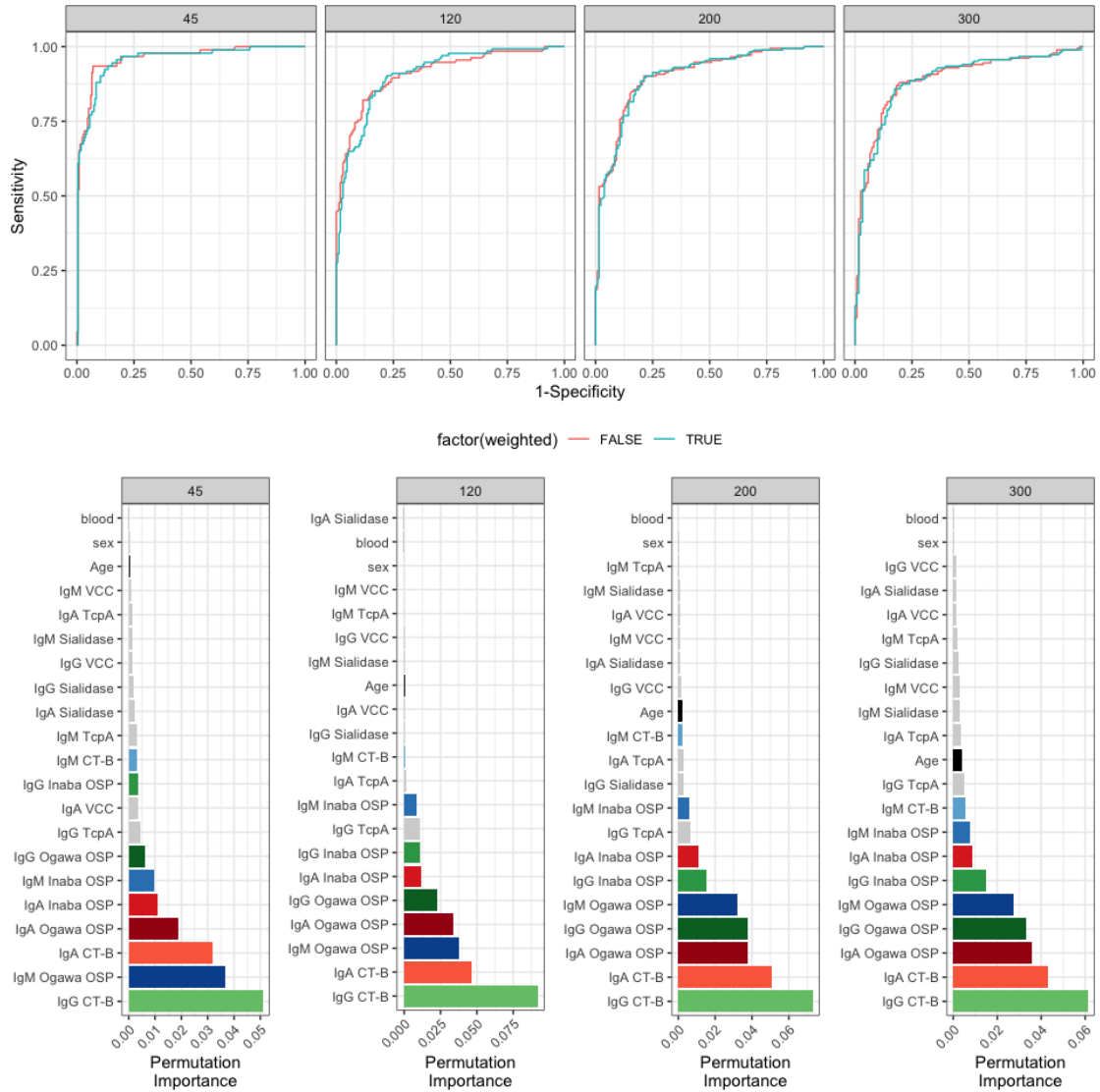


Marker	Covariate	Ratio of Baseline Values	Ratio of Fold-Change Values	Difference in days of Half-Life
	Under 10	1.25 (0.684-2.26)	1.01 (0.552-1.85)	-9 (-229-88)
MBA IgM Inaba	Female	2.91 (1.18-7.8)*	0.463 (0.145-1.27)	-13 (-29-4)
OSP	O Blood	0.97 (0.383-2.48)	1.17 (0.412-2.8)	19 (4-36)*
	Ogawa	0.34 (0.115-1.02)	0.616 (0.162-1.69)	19 (1-34)*
	Under 10	0.673 (0.269-1.71)	0.409 (0.12-1.19)	-17 (-32-0)
MBA IgM LTB	Female	1.46 (0.734-2.85)	0.681 (0.362-1.26)	0 (-18-158)
	O Blood	0.678 (0.342-1.29)	0.952 (0.503-1.74)	-4 (-129-33)
	Ogawa	0.43 (0.195-1.01)	0.717 (0.304-1.32)	-2 (-87-50)
	Under 10	1.21 (0.615-2.32)	0.902 (0.477-1.68)	-11 (-166-24)
MBA IgM LTh	Female	1.25 (0.649-2.3)	0.927 (0.431-1.92)	0 (-29-177)
	O Blood	0.69 (0.381-1.28)	0.988 (0.476-2.13)	-4 (-225-30)
	Ogawa	0.506 (0.224-1.08)	0.866 (0.327-1.76)	-2 (-118-106)
	Under 10	1.15 (0.618-2.12)	1.01 (0.483-2.16)	-6 (-239-18)
MBA IgM O139	Female	1.44 (0.806-2.48)	0.665 (0.455-0.979)*	-3 (-68-216)
	O Blood	0.809 (0.463-1.43)	0.885 (0.603-1.34)	-17 (-436-95)
	Ogawa	0.49 (0.255-0.978)*	0.802 (0.451-1.23)	0 (-243-317)
	Under 10	0.601 (0.349-1.01)	0.763 (0.524-1.13)	-56 (-368-35)
MBA IgM Ogawa	Female	1.49 (0.595-3.71)	0.312 (0.0998-0.939)*	9 (-22-51)
OSP	O Blood	1.48 (0.619-3.65)	0.867 (0.28-2.45)	-6 (-36-24)
	Ogawa	0.994 (0.334-3.14)	4.05 (1.19-10.9)*	35 (-9-61)
	Under 10	0.536 (0.224-1.29)	0.772 (0.239-2.28)	-6 (-35-25)
MBA IgM Sialidase	Female	2.06 (1.01-4.15)*	1.02 (0.438-2.51)	1 (-43-191)
	O Blood	1.17 (0.57-2.34)	0.709 (0.314-1.55)	-13 (-220-42)
	Ogawa	0.559 (0.231-1.32)	0.592 (0.168-1.35)	-1 (-179-87)
	Under 10	0.795 (0.409-1.56)	0.87 (0.38-1.9)	-6 (-159-84)
MBA IgM TcpA	Female	1.69 (0.679-4.29)	0.966 (0.386-1.7)	202 (-26-4953)
	O Blood	1.06 (0.455-2.68)	0.717 (0.425-1.17)	-151 (-817-32)
	Ogawa	0.629 (0.21-1.85)	0.776 (0.325-1.42)	123 (-49-790)
	Under 10	1.17 (0.493-2.85)	0.613 (0.377-1.02)	-204 (-757--2)*
	Female	1.61 (0.84-3.05)	1.1 (0.644-1.83)	1 (-57-90)

Marker	Covariate	Ratio of Baseline Values	Ratio of Fold-Change Values	Difference in days of Half-Life
MBA IgM VCC	O Blood	0.561 (0.282-1.05)	0.798 (0.462-1.3)	-8 (-81-76)
	Ogawa	0.449 (0.202-1)	0.586 (0.276-1.02)	13 (-32-185)
	Under 10	0.869 (0.465-1.64)	0.817 (0.463-1.34)	-8 (-116-42)
Vibriocida   Inaba	Female	3.46 (1.11-10.7)*	0.787 (0.225-2.56)	-35 (-89-22)
	O Blood	1.37 (0.46-4.39)	0.652 (0.195-1.99)	17 (-38-87)
	Ogawa	0.357 (0.0861-1.55)	0.295 (0.0591-1.32)	34 (-22-89)
	Under 10	0.938 (0.311-2.89)	0.505 (0.152-1.53)	19 (-33-91)
Vibriocida   Ogawa	Female	0.495 (0.176-1.42)	0.595 (0.147-2.13)	448 (231-807)*
	O Blood	1.35 (0.494-4.26)	0.846 (0.222-2.71)	-44 (-248-76)
	Ogawa	0.488 (0.147-1.67)	1.24 (0.287-4.06)	118 (17-268)*
	Under 10	0.22 (0.0815-0.596)*	1.46 (0.405-4.59)	240 (105-462)*

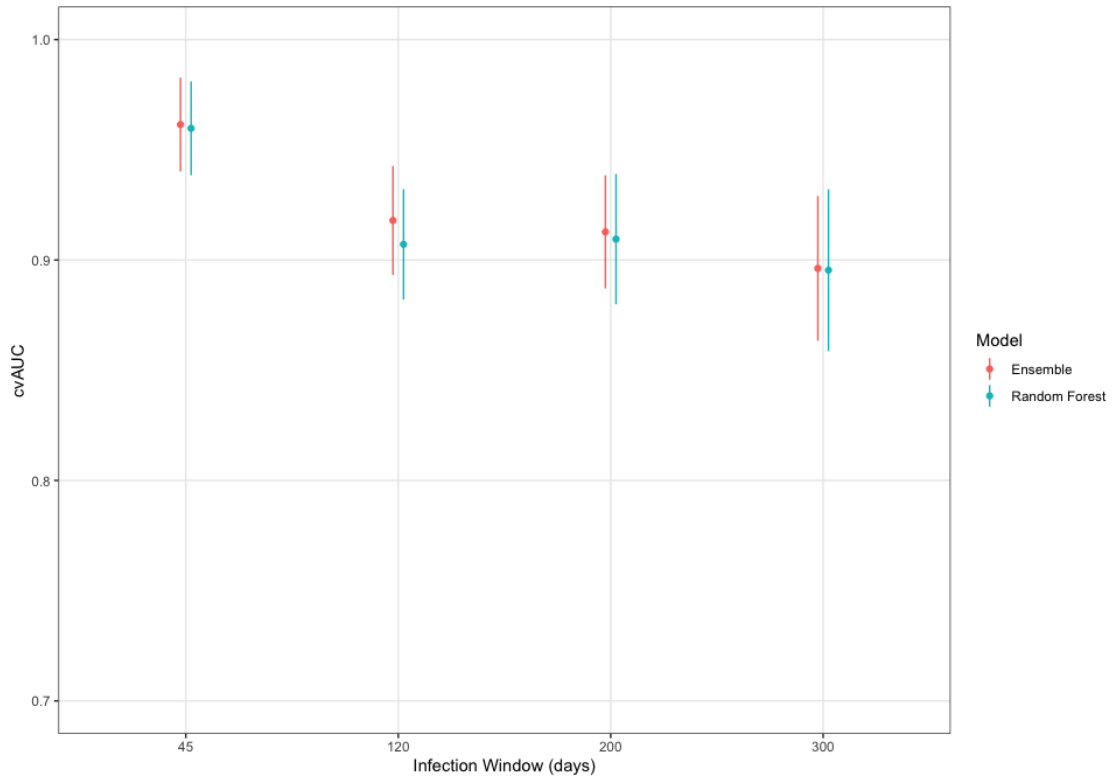
**Figure B.20. Cross-validated receiver operator characteristic curves and permutation importance for random forest models using 45-day, 120-day, 200-day, and 300-day infection window**

Permutation importance is shown along the x-axis for 21 predictors of random forest models containing multiplex bead assay (MBA) markers (except for those binding to CT-H, LT-H, and LT-B), age, sex, and blood-type.



### Figure B.21. Comparison of cross-validated area under the ROC curve between ensemble and random forest models

Models were fit to 18 MBA markers and 3 demographic variables (age, sex, and blood type). The ensemble model was created using the R package SuperLearner. Four different types of models were combined into the ensemble: Random Forest models (ranger), Lasso and Elastic-Net Regularized Generalized Linear Models (glmnet), Bayesian Additive Regression Trees (bartMachine), and Extreme Gradient Boosting (xgboost). All models were unweighted.



**Table B.5. Comparison of cross-validated AUC between multiple marker random forest models using 45-day, 120-day, 200-day, and 300-day infection windows**

Random forest models were fit using a the specified marker set and individual level factors including age, sex, and blood group. Mean and 95% confidence intervals for cvAUC are reported.

Marker Set	45-day	120-day	200-day	300-day
All MBA Markers	0.95 (0.93-0.97)	0.92 (0.89-0.94)	0.92 (0.89-0.94)	0.90 (0.87-0.94)
ELISA Markers	0.94 (0.91-0.98)	0.87 (0.84-0.91)	0.80 (0.75-0.85)	0.77 (0.72-0.83)
IgA MBA Markers	0.93 (0.90-0.96)	0.89 (0.86-0.92)	0.88 (0.86-0.91)	0.85 (0.81-0.88)
IgG MBA Markers	0.93 (0.90-0.96)	0.90 (0.87-0.92)	0.90 (0.87-0.93)	0.87 (0.82-0.91)
IgM MBA Markers	0.91 (0.87-0.95)	0.85 (0.81-0.88)	0.81 (0.76-0.85)	0.79 (0.74-0.84)
Vibriocidal Markers	0.92 (0.88-0.95)	0.86 (0.82-0.89)	0.82 (0.78-0.87)	0.84 (0.80-0.88)
Vibriocidal & ELISA Markers	0.97 (0.95-0.99)	0.92 (0.89-0.94)	0.87 (0.84-0.90)	0.88 (0.85-0.91)

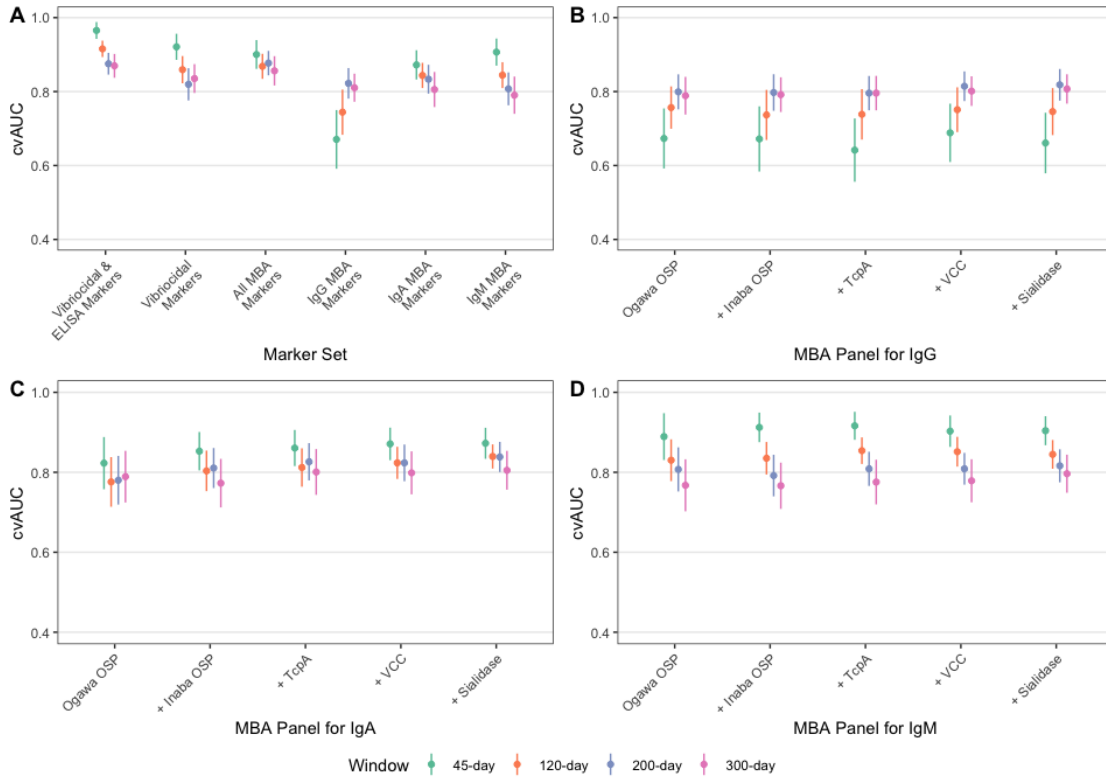
**Table B.6. Comparison of cross-validated AUC between multiplex bead assay IgG multiple marker random forest models using 45-day, 120-day, 200-day, and 300-day infection windows**

Random forest models were using a reduced panel of MBA IgG markers and individual level factors including age, sex, and blood group. Mean and 95% confidence intervals for cvAUC are reported.

MBA Panel for IgG	45-day	120-day	200-day	300-day
CT-B	0.90 (0.87-0.94)	0.88 (0.85-0.92)	0.85 (0.81-0.89)	0.82 (0.78-0.86)
+ Ogawa OSP	0.92 (0.89-0.96)	0.89 (0.86-0.92)	0.88 (0.85-0.92)	0.87 (0.83-0.91)
+ Inaba OSP	0.92 (0.89-0.96)	0.88 (0.85-0.92)	0.89 (0.86-0.92)	0.86 (0.82-0.90)
+ TcpA	0.93 (0.90-0.96)	0.90 (0.88-0.93)	0.90 (0.87-0.93)	0.88 (0.85-0.91)
+ VCC	0.92 (0.89-0.95)	0.89 (0.87-0.92)	0.89 (0.86-0.92)	0.87 (0.84-0.91)
+ Sialidase	0.93 (0.90-0.96)	0.90 (0.87-0.93)	0.90 (0.87-0.93)	0.87 (0.83-0.92)

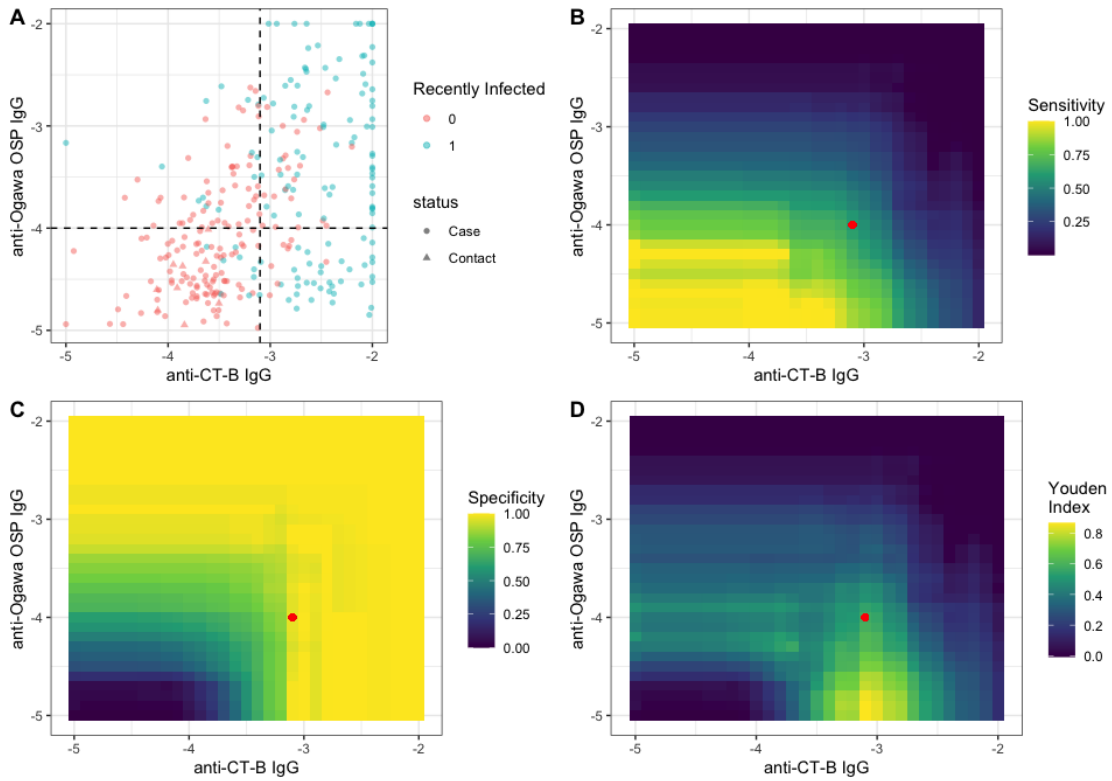
**Figure B.22. Comparison of cross-validated AUC across random forest models trained on traditional and MBA serological markers excluding anti-CT-B markers for 45-day, 120-day, 200-day, and 300-day infection windows**

Random forest models were fit using a the specified marker set and individual level factors including age, sex, and blood group (A). Estimated mean and 95% confidence intervals for cvAUC are reported. (B-D) Models fit to reduced panels of IgG, IgA, and IgM MBA markers are shown.



### Figure B.23. Marker cutoffs to detect recent infection within the past 120 days

Individuals were considered seropositive if their RAU was above the cut-off for both anti-Ogawa OSP IgG and anti-CT-B IgG. Sensitivity and specificity were estimated using a hierarchical logistic regression model with only an intercept fixed effect. The red dot represents the point where specificity is  $\geq 90\%$ , anti-OgawaOSP IgG RAU  $\geq -4.0$ , and then maximizing sensitivity.





## References

1. Ritz C, Baty F, Streibig JC, Gerhard D. Dose-response analysis using R. *PLoS one*. 2015 Dec 30;10(12):e0146021.

**Appendix C: Supplement to Conducting cholera serosurveillance in  
partially vaccinated populations**

**Method C.1. Equations for Bayesian framework to estimate seroincidence using known vaccination status**

*Known vaccination status adjustment*

We sought to estimate seroincidence (i.e.  $\lambda$ ) and vaccination coverage (i.e.  $\gamma$ ) in the context of a simulated serosurvey occurring shortly after a vaccination campaign using serological data and vaccination status. Vaccination status ( $V_i$ ) of individuals (indexed by  $i$ ) equals either 1 (vaccinated) or 0 (unvaccinated).  $S_i$  is the serostatus as determined by a classification model prediction from serological data. It can equal either 1 (seropositive) or 0 (seronegative).  $R_i$  is the true infection status, equaling either 1 (recently infected) or 0 (not recently infected). We used estimates of sensitivity and specificity for  $Pr(S_i|V_i, R_i)$ , including the estimated uncertainty in each parameter when only 3 IgG markers are measured. We assumed that the sensitivity of detecting recent infection was the same regardless of vaccination status (i.e.  $Pr(S_i|V_i = 1, R_i = 1) = Pr(S_i|V_i = 0, R_i = 1)$ ). Here is the likelihood function used in the framework:

$$\begin{aligned}
 Pr(\vec{S}, \vec{V}|\lambda, \gamma) &\approx \prod_{i=1}^n Pr(S_i, V_i|\lambda, \gamma) \\
 &= \prod_{i=1}^n \sum_{r=0}^1 Pr(S_i, V_i|R_i = r, \lambda, \gamma) \times Pr(R_i = r|\lambda, \gamma) \\
 &= \prod_{i=1}^n \sum_{r=0}^1 Pr(S_i|V_i, R_i = r, \lambda, \gamma) \times Pr(V_i|R_i = r, \lambda, \gamma) \times Pr(R_i = r|\lambda, \gamma) \\
 &= \prod_{i=1}^n \sum_{r=0}^1 Pr(S_i|V_i, R_i = r) \times Pr(V_i|\gamma) \times Pr(R_i = r|\lambda) \\
 &= \prod_{i=1}^n Pr(S_i|V_i, R_i = 0) \times Pr(V_i|\gamma) \times Pr(R_i = 0|\lambda) \\
 &\quad + Pr(S_i|V_i, R_i = 1) \times Pr(V_i|\gamma) \times Pr(R_i = 1|\lambda) \\
 &= \prod_{i=1}^n Pr(S_i|V_i, R_i = 0) \times Pr(V_i|\gamma) \times (1 - \lambda) \\
 &\quad + Pr(S_i|V_i, R_i = 1) \times Pr(V_i|\gamma) \times \lambda
 \end{aligned}$$

### Ignoring vaccination status

For the ignoring vaccination status “method”, we modified the equations for the known vaccination status adjustment by simply setting all individuals to be considered unvaccinated. Without any information about vaccination status, we do not estimate vaccination coverage,  $\gamma$ . We used estimates of sensitivity and specificity for  $Pr(S_i|V_i = 0, R_i)$ , including the estimated uncertainty in each parameter when only 3 IgG markers are measured. Here is the likelihood function used in the framework:

$$Pr(\vec{S}|\lambda) \approx \prod_{i=1}^n Pr(S_i|V_i = 0, R_i = 0) \times (1 - \lambda) \\ + Pr(S_i|V_i = 0, R_i = 1) \times \lambda$$

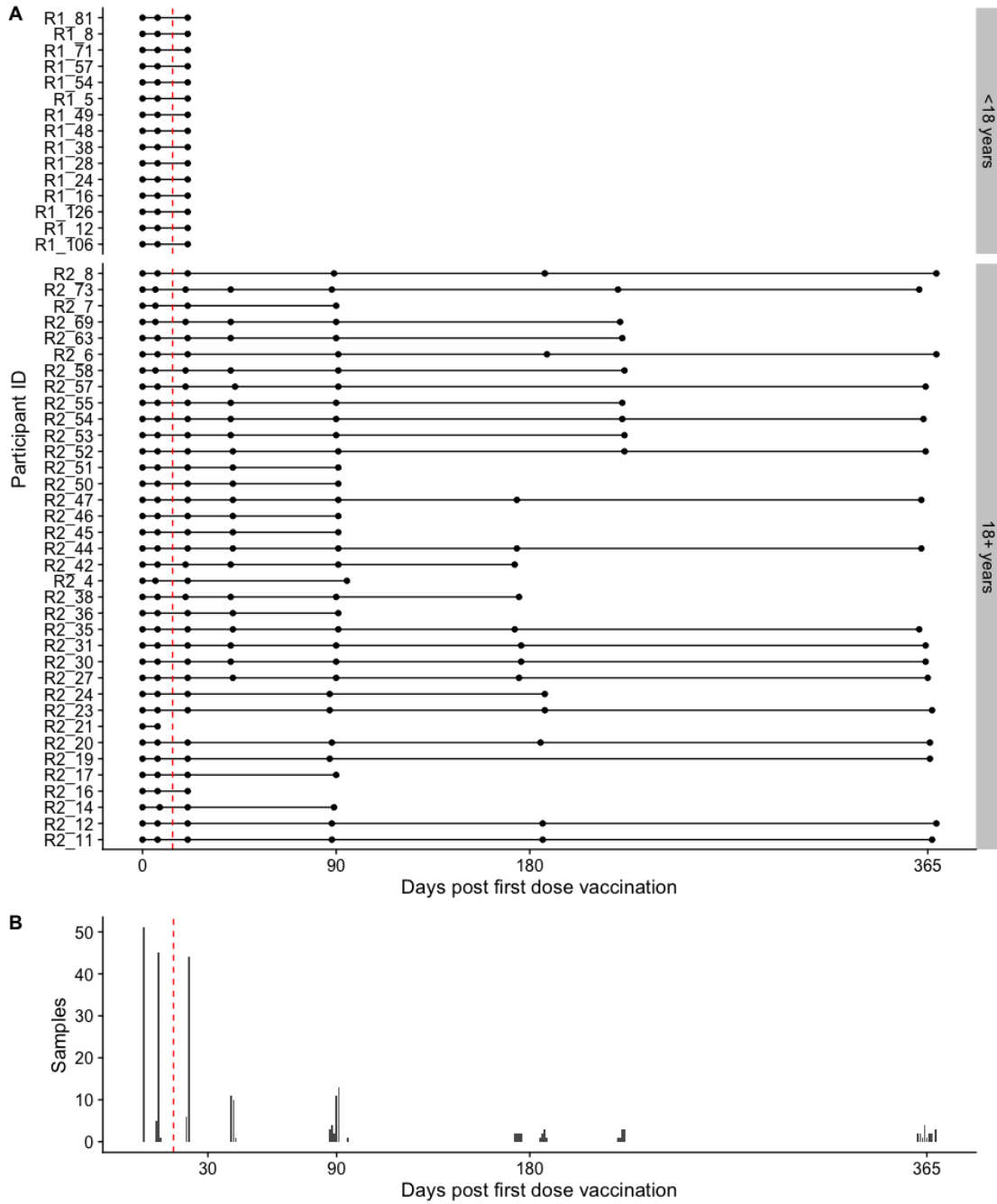
### **Method C.2. Equations for Bayesian framework to estimate seroincidence using additional serological data**

We sought to estimate seroincidence (i.e.  $\lambda$ ) and vaccination coverage (i.e.  $\gamma$ ).  $S_i$  is the serostatus as determined by a classification model using serological data from 15 markers. It can equal either 3 (seropositive, recently infected), 2 (seropositive, recently vaccinated but not recently infected), and 1 (seronegative).  $R_i$  is the true infection status, equaling either 1 (recently infected) or 0 (not recently infected).  $V_i$  is the true vaccination status, equaling either 1 (recently vaccinated) or 0 (not recently vaccinated). We used estimates of sensitivity and specificity for  $Pr(S_i|V_i, R_i)$ , including the estimated uncertainty in each parameter when 15 serological markers are measured. We assumed that the sensitivity of detecting recent infection was the same regardless of vaccination status (i.e.  $Pr(S_i|V_i = 1, R_i = 1) = Pr(S_i|V_i = 0, R_i = 1)$ ). Here is the likelihood used in the framework:

$$\begin{aligned}
Pr(\vec{S}|\lambda, \gamma) &\approx \prod_{i=1}^n Pr(S_i|\lambda, \gamma) \\
&= \prod_{i=1}^n \sum_{r=0}^1 \sum_{v=0}^1 Pr(S_i|R_i = r, V_i = v, \lambda, \gamma) \times Pr(R_i = r, V_i = v|\lambda, \gamma) \\
&= \prod_{i=1}^n \sum_{r=0}^1 \sum_{v=0}^1 Pr(S_i|R_i = r, V_i = v) \times Pr(R_i = r, V_i = v|\lambda, \gamma) \\
&= \prod_{i=1}^n \sum_{r=0}^1 \sum_{v=0}^1 Pr(S_i|R_i = r, V_i = v) \times Pr(R_i = r|\lambda) \times Pr(V_i = v|\gamma) \\
&= \prod_{i=1}^n Pr(S_i|R_i = 1, V_i = 1) \times Pr(R_i = 1|\lambda) \times Pr(V_i = 1|\gamma) \\
&\quad + Pr(S_i|R_i = 1, V_i = 0) \times Pr(R_i = 1|\lambda) \times Pr(V_i = 0|\gamma) \\
&\quad + Pr(S_i|R_i = 0, V_i = 1) \times Pr(R_i = 0|\lambda) \times Pr(V_i = 1|\gamma) \\
&\quad + Pr(S_i|R_i = 0, V_i = 0) \times Pr(R_i = 0|\lambda) \times Pr(V_i = 0|\gamma) \\
&= \prod_{i=1}^n Pr(S_i|R_i = 1, V_i = 1) \times \lambda\gamma \\
&\quad + Pr(S_i|R_i = 1, V_i = 0) \times \lambda(1 - \gamma) \\
&\quad + Pr(S_i|R_i = 0, V_i = 1) \times (1 - \lambda)\gamma \\
&\quad + Pr(S_i|R_i = 0, V_i = 0) \times (1 - \lambda)(1 - \gamma) \\
&\quad \text{Assume } S_i|R_i = 1 \text{ does not depend } V_i \\
&= \prod_{i=1}^n Pr(S_i|R_i = 1) \times \lambda \\
&\quad + Pr(S_i|R_i = 0, V_i = 1) \times (1 - \lambda)\gamma \\
&\quad + Pr(S_i|R_i = 0, V_i = 0) \times (1 - \lambda)(1 - \gamma)
\end{aligned}$$

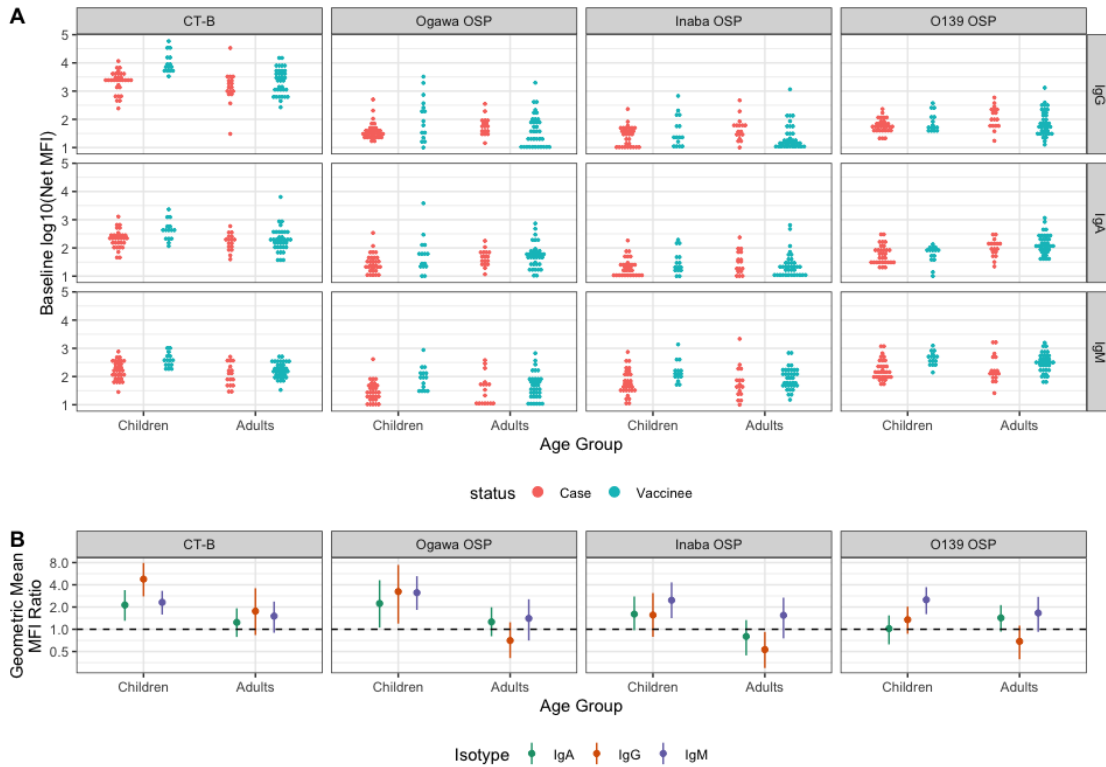
**Figure C.1. Timing of sample collection among Haitian vaccinees relative to date of first dose vaccination**

Each point represents a blood sample collection occurred for a vaccinated volunteer (A). The daily number of samples is shown on the y-axis (B). Red-dashed line indicates timing of second dose at 14 days.



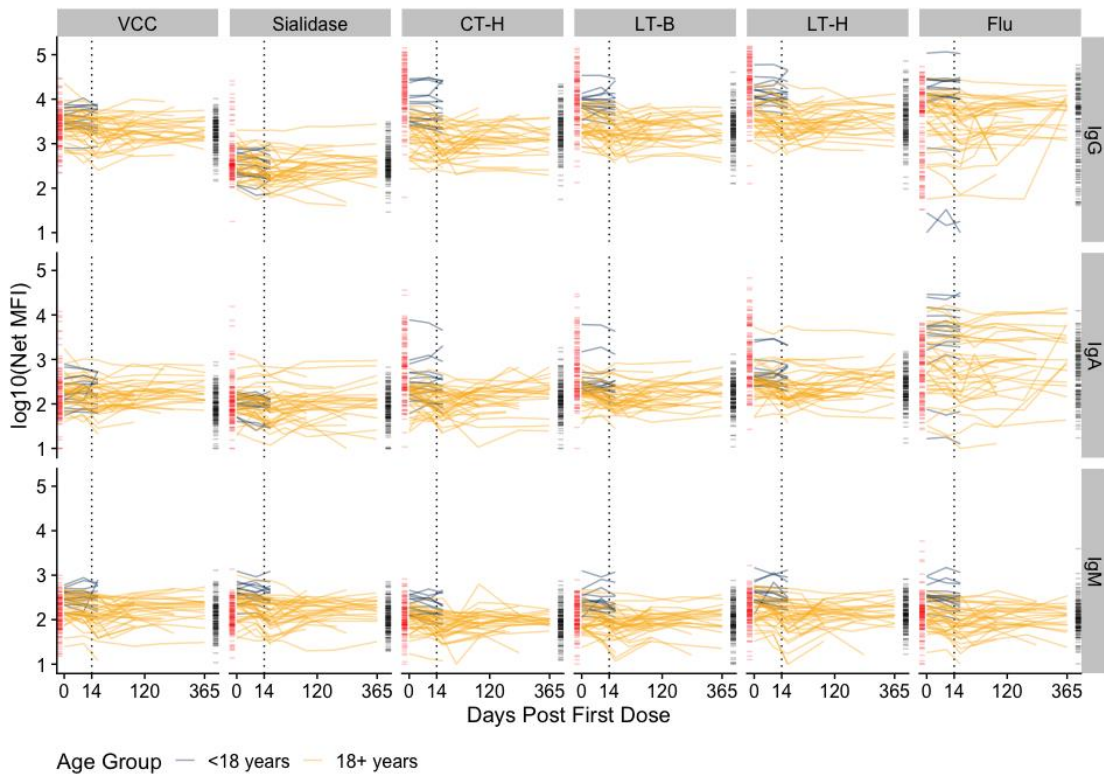
**Figure C.2. Comparison of baseline antibody measurements among Haitian vaccinees and Bangladeshi cholera cases**

Baseline antibody measurements from cases (2-4 days after infection) and vaccinees (day of first dose) are shown on the y-axis (A). Ratios of the geometric mean MFI are shown on the y-axis with 95% bootstrapped confidence intervals indicated by vertical lines (B).



**Figure C.3. Multiplex bead assay net median fluorescence intensity measurements of IgG, IgM, and IgA against other antigens among Haitian vaccinated volunteers**

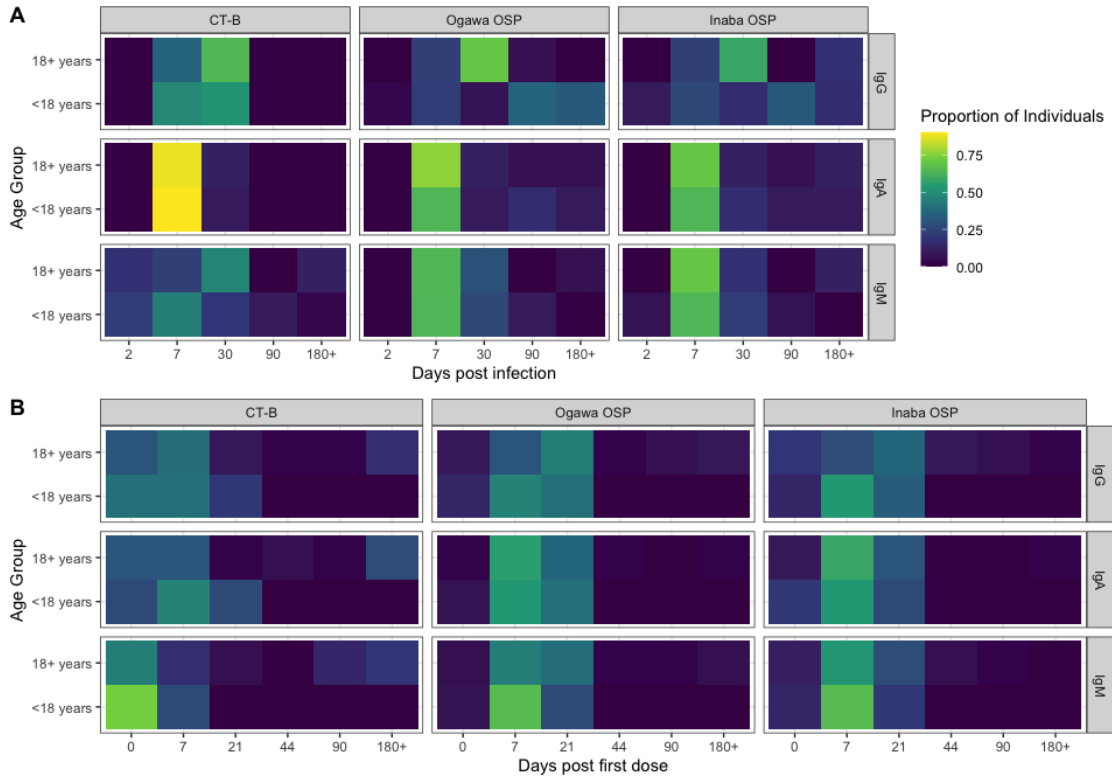
Y-axis indicates the log (base 10) of the Net MFI. X-axis is square-root transformed. Each colored line indicates individual trajectories over time (dark blue: children, gold: adults). Rug plots show the antibody measurements from the cohort in Bangladesh (red: Case measurements inside the 120-day infection window; black: Uninfected household contacts and case measurements outside a 120-day infection window). The black dotted line indicates the timing of second dose vaccination.





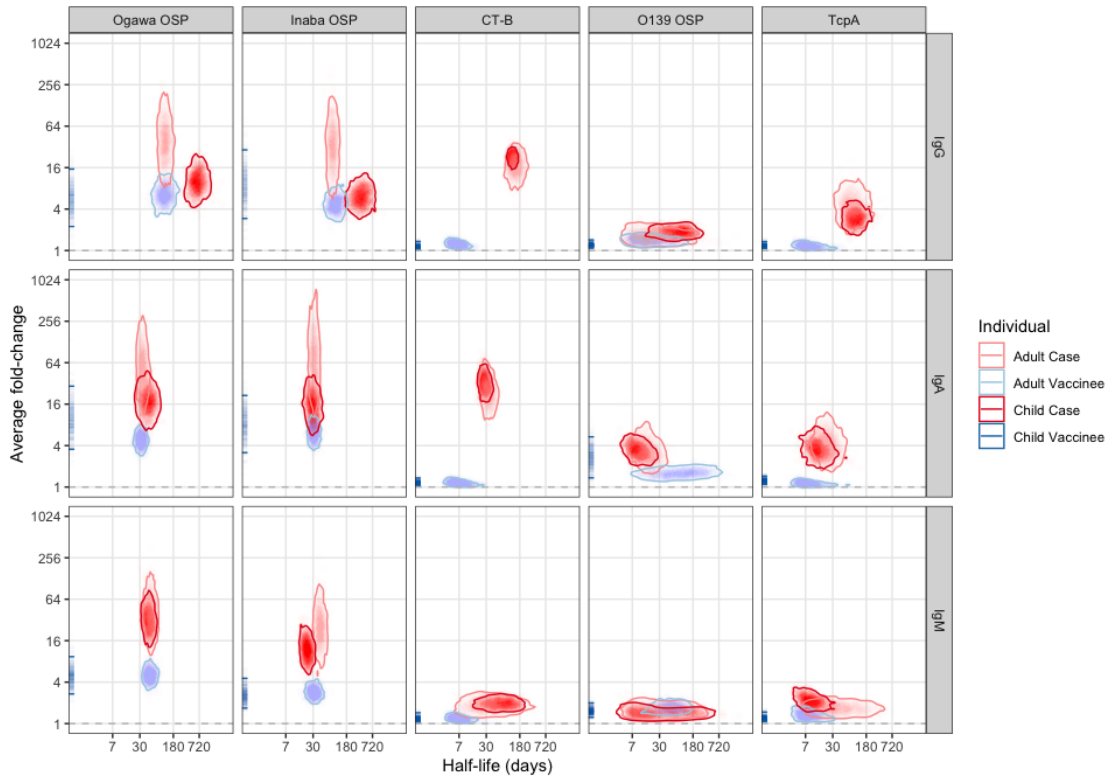
**Figure C.4. Proportion of individuals with a peak value on a given day, by age group and antibody**

Color indicates the proportion of cases(A) and vaccinees (B) whose peak value fall on around a certain day since infection or first dose vaccination. Proportions are stratified by age group and antibody.



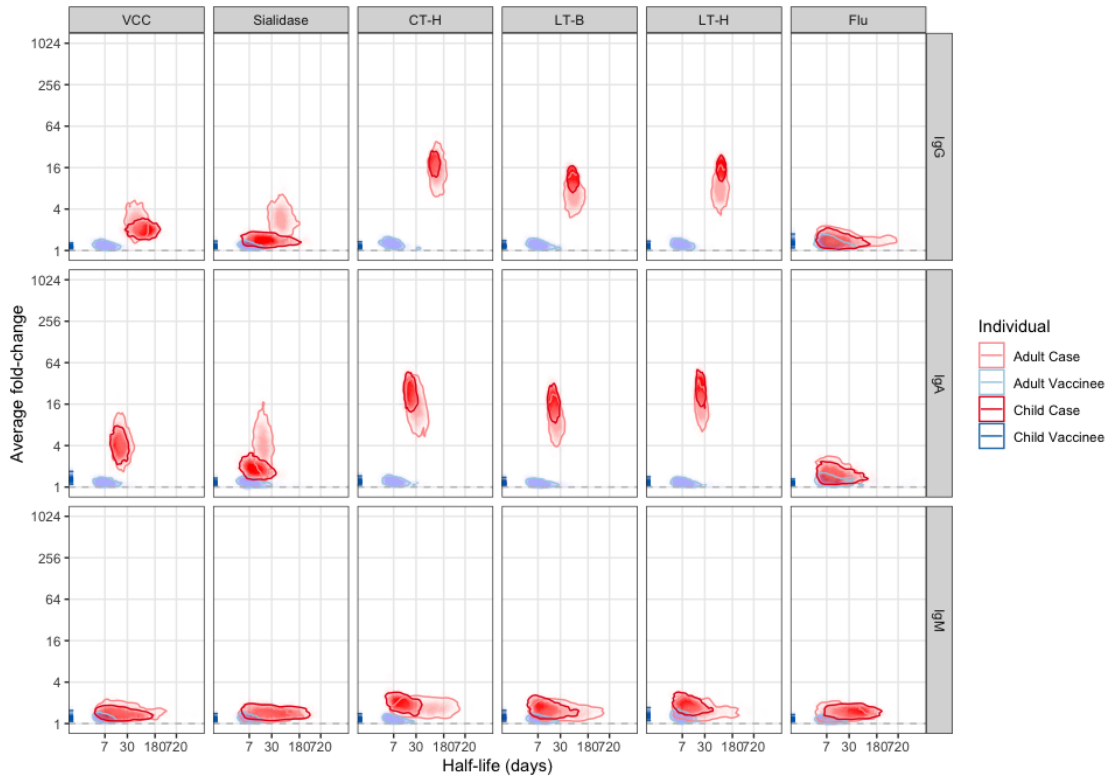
**Figure C.5. Posterior density of half-life and average fold-change from exponential kinetic models for anti-OSP, anti-CT-B, and anti-TcpA antibodies among vaccinees and cases**

Points represent 1000 draws from the posterior distribution for the average individual (red: children, pink: adults) and vaccinees (light blue: adults) and solid lines highlighting a contour containing the top 95% of all estimates. Estimates of the average fold change for child vaccinees are shown as a rug plot (dark blue) with solid lines indicating the 95% credible interval.



**Figure C.6. Posterior density of half-life and average fold-change from exponential kinetic models for additional antibody measurements among vaccinees and cases**

Points represent 1000 draws from the posterior distribution for the average individual (red: children, pink: adults) and vaccinees (light blue: adults) and solid lines highlighting a contour containing the top 95% of all estimates. Estimates of the average fold change for child vaccinees are shown as a rug plot (dark blue) with solid lines indicating the 95% credible interval.



**Table C.1. Estimated duration of half-life and average fold-change from exponential kinetic models for vaccinees and cases**

The median estimate duration of half-life (in days) and the median estimate increase in fold-change for the average individual are shown below. 95% Bayesian Credible Intervals are shown in parentheses.

Isotype	Antigen	Status	Average Fold-Change (95% CI)	Half Life (95% CI) in days
IgG	CT-B	Adult Case	16.5 (9.1-30.9)	145 (94-238)
		Adult Vaccinee	1.2 (1.1-1.4)	6 (4-12)
		Child Case	22.2 (16.6-29.5)	123 (98-156)
		Child Vaccinee	1.1 (1-1.4)	10 (5-73)
	Ogawa OSP	Adult Case	38.4 (11.3-134.5)	116 (84-168)
		Adult Vaccinee	6.4 (3.9-11.1)	114 (67-193)
		Child Case	9.9 (5.2-19.7)	641 (389-1153)
		Child Vaccinee	5.5 (2.2-15.2)	156 (63-1635)
	Inaba OSP	Adult Case	33.9 (9.2-121.2)	86 (65-113)
		Adult Vaccinee	4.6 (3-7.6)	101 (62-163)
		Child Case	6 (3.5-10.4)	392 (201-715)
		Child Vaccinee	8.5 (2.9-28.9)	179 (62-3296)
	O139 OSP	Adult Case	1.6 (1.2-2.3)	19 (6-184)
		Adult Vaccinee	1.4 (1.2-1.7)	14 (6-98)
		Child Case	1.9 (1.5-2.4)	83 (20-218)
		Child Vaccinee	1.2 (1.1-1.4)	15 (5-147)
	TcpA	Adult Case	4.4 (2.4-8.8)	82 (42-194)
		Adult Vaccinee	1.2 (1.1-1.4)	8 (4-29)
		Child Case	3 (1.9-4.6)	100 (57-190)
		Child Vaccinee	1.2 (1.1-1.3)	12 (5-94)
VCC	Adult Case	2.5 (1.5-4.5)	53 (31-97)	
	Adult Vaccinee	1.2 (1.1-1.4)	7 (4-16)	
	Child Case	2 (1.6-2.6)	89 (35-186)	
	Child Vaccinee	1.1 (1-1.3)	10 (5-50)	
Sialidase	Adult Case	2.9 (1.7-5.4)	55 (29-111)	

Isotype	Antigen	Status	Average Fold-Change (95% CI)	Half Life (95% CI) in days
		Adult Vaccinee	1.1 (1.1-1.3)	7 (4-16)
		Child Case	1.4 (1.2-1.7)	19 (6-122)
		Child Vaccinee	1.2 (1-1.4)	9 (4-52)
	CT-H	Adult Case	14.7 (7.3-28.6)	116 (76-187)
		Adult Vaccinee	1.3 (1.1-1.5)	6 (3-14)
		Child Case	18.2 (12.6-25.3)	99 (76-131)
		Child Vaccinee	1.1 (1-1.3)	11 (5-129)
	LT-B	Adult Case	6.4 (3.7-11.1)	70 (43-121)
		Adult Vaccinee	1.2 (1.1-1.4)	7 (4-20)
		Child Case	10.9 (7.7-15.6)	67 (50-90)
		Child Vaccinee	1.2 (1-1.4)	11 (5-94)
	LT-H	Adult Case	7.7 (4.4-13.6)	80 (53-123)
		Adult Vaccinee	1.2 (1.1-1.4)	6 (4-14)
		Child Case	15.3 (11.2-21.4)	87 (70-110)
		Child Vaccinee	1.2 (1.1-1.4)	10 (5-85)
	Flu	Adult Case	1.4 (1.1-2)	13 (5-427)
		Adult Vaccinee	1.3 (1.1-1.6)	8 (4-36)
		Child Case	1.4 (1.1-1.8)	12 (5-74)
		Child Vaccinee	1.3 (1.1-1.7)	8 (4-36)
IgA	CT-B	Adult Case	24.7 (11.8-53.7)	33 (23-51)
		Adult Vaccinee	1.1 (1-1.3)	6 (4-19)
		Child Case	31.6 (19.6-54)	27 (21-38)
		Child Vaccinee	1.2 (1-1.4)	10 (5-228)
	Ogawa OSP	Adult Case	55.4 (18.3-205)	35 (27-46)
		Adult Vaccinee	4.8 (3.2-7.2)	32 (23-44)
		Child Case	17.6 (8.5-36.8)	46 (26-78)
		Child Vaccinee	9.6 (3.5-29.2)	43 (19-250)
	Inaba OSP	Adult Case	74.9 (18.7-377.1)	32 (25-41)
		Adult Vaccinee	6.2 (3.9-9.7)	32 (24-42)

Isotype	Antigen	Status	Average Fold-Change (95% CI)	Half Life (95% CI) in days
		Child Case	15.8 (7.3-33.3)	29 (20-45)
		Child Vaccinee	8.2 (3.2-21.4)	42 (16-401)
	O139 OSP	Adult Case	3.3 (1.9-6.7)	15 (7-36)
		Adult Vaccinee	1.6 (1.3-2)	93 (11-525)
		Child Case	3.4 (2.3-5.2)	9 (5-21)
		Child Vaccinee	2.7 (1.4-5.4)	119 (30-2612)
	TcpA	Adult Case	4.2 (2.1-9.1)	24 (12-46)
		Adult Vaccinee	1.1 (1-1.3)	7 (4-29)
		Child Case	3.5 (2.2-6.2)	13 (7-31)
		Child Vaccinee	1.2 (1.1-1.4)	10 (5-43)
	VCC	Adult Case	4.4 (2.2-9.6)	21 (13-36)
		Adult Vaccinee	1.2 (1.1-1.3)	7 (4-17)
		Child Case	4.1 (2.5-6.7)	18 (11-28)
		Child Vaccinee	1.3 (1.1-1.7)	14 (5-206)
	Sialidase	Adult Case	4 (1.7-10.3)	17 (11-27)
		Adult Vaccinee	1.1 (1-1.3)	7 (4-26)
		Child Case	1.8 (1.4-2.6)	10 (5-29)
		Child Vaccinee	1.2 (1.1-1.3)	10 (5-61)
	CT-H	Adult Case	15.1 (6.8-33.7)	31 (18-53)
		Adult Vaccinee	1.2 (1.1-1.4)	7 (4-18)
		Child Case	23.6 (14.4-39.7)	21 (15-29)
		Child Vaccinee	1.2 (1.1-1.5)	9 (5-47)
	LT-B	Adult Case	9.3 (4.6-17.7)	22 (15-34)
		Adult Vaccinee	1.1 (1-1.3)	7 (4-24)
		Child Case	16.3 (9.8-26.9)	19 (15-26)
		Child Vaccinee	1.2 (1-1.4)	9 (4-37)
	LT-H	Adult Case	15 (8.1-27.1)	24 (17-34)
		Adult Vaccinee	1.1 (1-1.3)	7 (4-20)
		Child Case	26.9 (16.7-41.9)	23 (18-30)

Isotype	Antigen	Status	Average Fold-Change (95% CI)	Half Life (95% CI) in days
IgM	Flu	Child Vaccinee	1.2 (1.1-1.4)	10 (4-61)
		Adult Case	1.6 (1.2-2.5)	11 (5-51)
		Adult Vaccinee	1.2 (1.1-1.5)	8 (4-58)
		Child Case	1.5 (1.2-2.1)	11 (5-60)
		Child Vaccinee	1.1 (1.1-1.3)	10 (5-61)
	CT-B	Adult Case	1.8 (1.3-2.6)	34 (8-210)
		Adult Vaccinee	1.2 (1.1-1.4)	7 (4-17)
		Child Case	1.9 (1.6-2.5)	68 (19-177)
		Child Vaccinee	1.2 (1-1.4)	8 (4-32)
	Ogawa OSP	Adult Case	38.6 (14.5-115.2)	54 (39-74)
		Adult Vaccinee	5 (3.4-7.5)	51 (37-75)
		Child Case	32 (16.4-67)	48 (35-67)
		Child Vaccinee	4.9 (2.7-9.4)	58 (19-866)
	Inaba OSP	Adult Case	24.7 (9.1-75.2)	44 (33-59)
		Adult Vaccinee	2.8 (2.1-3.9)	32 (21-46)
		Child Case	11.8 (6.2-22.1)	22 (16-31)
		Child Vaccinee	2.6 (1.7-4.5)	29 (9-424)
	O139 OSP	Adult Case	1.5 (1.2-2.2)	28 (6-238)
		Adult Vaccinee	1.7 (1.4-2.1)	55 (18-133)
		Child Case	1.4 (1.2-1.8)	19 (5-331)
		Child Vaccinee	1.5 (1.2-2)	20 (7-105)
	TcpA	Adult Case	1.7 (1.3-2.2)	37 (10-270)
		Adult Vaccinee	1.3 (1.1-1.6)	7 (4-27)
		Child Case	2.1 (1.6-3)	9 (4-22)
Child Vaccinee		1.2 (1.1-1.5)	8 (4-30)	
VCC	Adult Case	1.5 (1.2-2)	19 (6-210)	
	Adult Vaccinee	1.2 (1.1-1.4)	6 (4-14)	
	Child Case	1.4 (1.2-1.7)	16 (5-100)	
	Child Vaccinee	1.2 (1.1-1.5)	14 (5-131)	

Isotype	Antigen	Status	Average Fold-Change (95% CI)	Half Life (95% CI) in days
Sialidase		Adult Case	1.3 (1.1-1.8)	16 (5-146)
		Adult Vaccinee	1.2 (1.1-1.4)	6 (4-15)
		Child Case	1.4 (1.2-1.8)	29 (6-225)
		Child Vaccinee	1.2 (1.1-1.4)	9 (4-51)
CT-H		Adult Case	1.7 (1.3-2.4)	58 (9-299)
		Adult Vaccinee	1.2 (1.1-1.3)	7 (4-19)
		Child Case	2 (1.5-2.6)	12 (5-34)
		Child Vaccinee	1.2 (1-1.4)	8 (4-39)
LT-B		Adult Case	1.5 (1.2-2.1)	24 (6-272)
		Adult Vaccinee	1.1 (1-1.3)	7 (4-21)
		Child Case	1.6 (1.3-2.2)	11 (5-69)
		Child Vaccinee	1.2 (1.1-1.6)	8 (4-47)
LT-H		Adult Case	1.5 (1.2-2.2)	17 (5-174)
		Adult Vaccinee	1.2 (1.1-1.5)	7 (4-15)
		Child Case	1.8 (1.4-2.6)	10 (5-32)
		Child Vaccinee	1.3 (1.1-1.7)	9 (4-40)
Flu		Adult Case	1.4 (1.1-1.9)	14 (5-101)
		Adult Vaccinee	1.1 (1-1.3)	7 (4-22)
		Child Case	1.5 (1.2-1.8)	39 (8-156)
		Child Vaccinee	1.1 (1.1-1.4)	8 (4-31)



**Table C.2. Number of samples from vaccinated individuals misclassified as recently infected by infection window**

Age Group	Days after first dose	45-day model	120-day model	200-day model	300-day model
<18 years	0	7% (1/15)	20% (3/15)	27% (4/15)	27% (4/15)
	7	20% (3/15)	47% (7/15)	53% (8/15)	60% (9/15)
	21	27% (4/15)	33% (5/15)	33% (5/15)	47% (7/15)
18+ years	0	0% (0/36)	0% (0/36)	3% (1/36)	3% (1/36)
	7	0% (0/36)	3% (1/36)	17% (6/36)	17% (6/36)
	21	0% (0/35)	0% (0/35)	26% (9/35)	23% (8/35)
	44	0% (0/22)	0% (0/22)	14% (3/22)	14% (3/22)
	90	0% (0/34)	0% (0/34)	9% (3/34)	9% (3/34)
	180	0% (0/15)	0% (0/15)	7% (1/15)	7% (1/15)
	220	0% (0/8)	0% (0/8)	0% (0/8)	0% (0/8)
360	0% (0/17)	0% (0/17)	0% (0/17)	0% (0/17)	

**Table C.3. Average and range of seroincidence estimates in simulated serological surveys**

True Coverage	Days since second dose	Ignore	Known vaccination status	Serological
25%	7	17.5% (15.2%-19.8%)	6.6% (4.8%-8.5%)	6.4% (5.4%-8.1%)
	30	14.7% (11.4%-16.3%)	6.0% (4.6%-7.4%)	5.9% (5.0%-7.3%)
	76	9.9% (8.6%-10.9%)	6.2% (5.1%-7.6%)	6.0% (4.6%-8.4%)
50%	7	29.7% (25.3%-31.9%)	5.7% (3.9%-7.4%)	5.7% (4.2%-7.1%)
	30	25.1% (21.8%-29.4%)	6.5% (4.8%-9.4%)	6.0% (4.1%-8.1%)
	76	14.8% (10.7%-18.6%)	6.4% (4.9%-9.8%)	6.0% (5.1%-7.5%)
75%	7	43.5% (40.5%-49.3%)	7.5% (5.7%-9.7%)	6.6% (5.0%-7.8%)
	30	33.8% (30.7%-40.3%)	6.6% (4.0%-10.8%)	5.5% (4.4%-7.5%)
	76	19.0% (17.3%-22.6%)	6.3% (3.6%-10.4%)	6.0% (5.0%-7.6%)

**Table C.4. Average and range of coverage estimates in simulated serological surveys**

True Coverage	Days since second dose	Known Vaccination Status	Serological
25%	7	24.9% (22.5%-27.0%)	25.4% (22.7%-27.9%)
	30	25.4% (23.7%-27.4%)	26.6% (23.6%-29.6%)
	76	25.0% (23.8%-26.1%)	26.5% (22.2%-32.7%)
50%	7	49.9% (49.0%-50.9%)	50.1% (47.6%-53.6%)
	30	50.1% (48.6%-51.9%)	53.8% (52.2%-55.0%)
	76	49.3% (48.4%-50.3%)	52.2% (43.9%-56.6%)
75%	7	74.4% (73.0%-75.8%)	74.9% (72.6%-77.1%)
	30	75.0% (74.2%-77.3%)	75.6% (70.2%-81.2%)
	76	75.3% (74.6%-76.7%)	75.5% (69.7%-80.2%)

# CURRICULUM VITAE

## Areas of Interest

Epidemiology, serological surveillance, global health

## Education

- 8/2017 – PhD Candidate in Epidemiology (Infectious Disease Track)  
5/2022 Johns Hopkins Bloomberg School of Public Health, Baltimore, MD  
(expected) Advisors: Andrew Azman and Justin Lessler
- 4/2018 – MHS Student in Biostatistics  
5/2022 Johns Hopkins Bloomberg School of Public Health, Baltimore, MD  
(expected) Advisor: Elizabeth Colantuoni
- 9/2013 – MPH Epidemiology of Microbial Diseases (Concentration in Global Health)  
5/2015 Yale School of Public Health, New Haven, CT  
Advisors: Albert Ko and Alison Galvani
- 9/2009 – BS Chemical Biology (Honors, GPA 3.75/4.00)  
5/2013 University of California, Berkeley, College of Chemistry, Berkeley, CA

## Professional Experience

- 8/2017 - Graduate Student Researcher, [Infectious Disease Dynamics Group](#)  
5/2022 Supervisors: Andrew Azman, Justin Lessler, and Amy Wesolowski  
(expected) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
- 9/2018 - Student Coordinator, [Surveillance and Outbreak Response Team \(SORT\)](#)  
9/2021 Faculty Coordinators: Emily Gurley and Melissa Marx  
Johns Hopkins Bloomberg School of Public Health, Maryland Department of Health & Baltimore City Department of Health
- 9/2020 - COVID-19 Surveillance Team at the University of New Hampshire  
6/2021 Supervisors: Emily Gurley and David Celentano  
Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
- 6/2018 – Virology Stagiaire (Internship in the Virology Unit)  
7/2018 Institut Pasteur de Madagascar, Antananarivo, Madagascar
- 9/2015 – CDC/ASPPH Allan Rosenfield Global Epidemiology Fellowship  
8/2017 CDC Laboratory Branch / Tanzanian Field Epidemiology & Laboratory Training Program  
US Centers for Disease Control and Prevention, Dar es Salaam, Tanzania
- 11/2013 – Graduate Student Researcher, Center for Infectious Disease Modeling and Analysis  
9/2015 Yale School of Public Health, Yale University, New Haven, CT
- 5/2014 – World Health Organization Outbreak Investigation and Surveillance Internship  
8/2014 World Health Organization Country Office, Honiara, Solomon Islands

## Publications and Manuscripts

- [1] **Jones FK**, Bhuiyan TR, Slater D, Vater KRH, Khan AI, Tennier R, Visieres K, Ivers L, Chowdhury Fahima, Kelly M, Xu P, Kováč P, Biswas R, Kamruzzaman M, Ryan ET, Calderwood SB, LaRocque RC, Lessler J, Charles RC, Leung DT, Qadri F, Harris JB, Azman AS. Conducting cholera serosurveillance in partially vaccinated populations. (*Manuscript in preparation*)
- [2] **Jones FK**, Bhuiyan TR, Mills R, Khan AI, Slater D, Vater KRH, Chowdhury Fahima, Kelly M, Xu P, Kováč P, Biswas R, Kamruzzaman M, Ryan ET, Calderwood SB, LaRocque RC, Lessler J,

Charles RC, Leung DT, Qadri F, Harris JB, Azman AS. Identifying recent cholera infections using a multiplex bead serological assay. (*Manuscript submitted to eLife*)

- [3] **Jones FK**, Mensah K, Heraud JM, Randriatsarafara FM, Metcalf CJ, Wesolowski A. The challenge of achieving immunity through multiple-dose vaccines in Madagascar. *American Journal of Epidemiology*. 2021 May 22. doi: [10.1093/aje/kwab145](https://doi.org/10.1093/aje/kwab145)
- [4] Wiens KE, Mawien PN, Rumunu J, Slater D, **Jones FK**, Moheed S, Cafilisch A, Bior BK, Jacob IA, Lako RL, Guyo AG. Seroprevalence of Severe Acute Respiratory Syndrome Coronavirus 2 IgG in Juba, South Sudan, 2020. *Emerging infectious diseases*. 2021 Jun;27(6):1598. doi: [10.3201/eid2706.210568](https://doi.org/10.3201/eid2706.210568)
- [5] Razafimahatratra SL\*, Wesolowski A\*, Rafetrarivony L, Heraud JM, **Jones FK**, Cauchemez S, Razafindratsimandresy R, Raharinantoanina SJ, Harimanana A, Collard JM, Metcalf CJ. Seroprevalence of pertussis in Madagascar and implications for vaccination. *Epidemiology & Infection*. 2020;148. doi: [10.1017/S0950268820002800](https://doi.org/10.1017/S0950268820002800)
- [6] **Jones FK**, Wamala JF, Rumunu J, Mawien PN, Kol MT, Wohl S, Deng L, Pezzoli L, Omar LH, Lessler J, Quilici ML. Successive epidemic waves of cholera in South Sudan between 2014 and 2017: a descriptive epidemiological study. *The Lancet Planetary Health*. 2020 Dec 1;4(12):e577-87. doi: [10.1016/S2542-5196\(20\)30255-2](https://doi.org/10.1016/S2542-5196(20)30255-2)
- [7] Iyer AS\*, **Jones FK\***, Nodoushani A\*, Kelly M, Becker M, Slater D, Mills R, Teng E, Kamruzzaman M, Garcia-Beltran WF, Astudillo M. Persistence and decay of human antibody responses to the receptor binding domain of SARS-CoV-2 spike protein in COVID-19 patients. *Science immunology*. 2020 Oct 8;5(52). doi: [10.1126/sciimmunol.abe0367](https://doi.org/10.1126/sciimmunol.abe0367)
- [8] Zheng Q\*, **Jones FK\***, Leavitt SV\*, Ung L\*, Labrique AB, Peters DH, Lee EC, Azman AS. HIT-COVID, a global database tracking public health interventions to COVID-19. *Scientific data*. 2020 Aug 27;7(1):1-8. doi: [10.1038/s41597-020-00610-2](https://doi.org/10.1038/s41597-020-00610-2)
- [9] Lauer SA\*, Grantz KH\*, Bi Q, **Jones FK**, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*. 2020 May 5;172(9):577-82. doi: [10.7326/M20-0504](https://doi.org/10.7326/M20-0504)
- [10] Yamin D\*, **Jones FK\***, DeVincenzo JP, Gertler S, Kobiler O, Townsend JP, Galvani AP. Vaccination strategies against respiratory syncytial virus. *Proceedings of the National Academy of Sciences*. 2016 Nov 15;113(46):13239-44. doi: [10.1073/pnas.1522597113](https://doi.org/10.1073/pnas.1522597113)
- [11] **Jones FK**, Ko AI, Becha C, Joshua C, Musto J, Thomas S, Ronsse A, Kirkwood CD, Sio A, Aumua A, Nilles EJ. Increased Rotavirus Prevalence in Diarrheal Outbreak Precipitated by Localized Flooding, Solomon Islands, 2014. *Emerging infectious diseases*. 2016 May;22(5):875. doi: [10.3201/eid2205.151743](https://doi.org/10.3201/eid2205.151743)
- [12] Durham DP, Ndeffo-Mbah ML, Skrip LA, **Jones FK**, Bauch CT, Galvani AP. National-and state-level impact and cost-effectiveness of nonavalent HPV vaccination in the United States. *Proceedings of the National Academy of Sciences*. 2016 May 3;113(18):5107-12. doi: [10.1073/pnas.1515528113](https://doi.org/10.1073/pnas.1515528113)

**\*Contributed equally**

**Oral Presentations**

- [1] Dissertation Defense Seminar, Johns Hopkins Bloomberg School of Public Health. 2022 March 17
- [2] Current Topics in Epidemiology, Johns Hopkins Bloomberg School of Public Health. 2019 December 13
- [3] Biostatistics Departmental Seminar, Johns Hopkins Bloomberg School of Public Health. 2019 September 9

- [4] Seminar at Faculty of Engineering, Tel Aviv University. 2017 January 19.
- [5] African Society for Laboratory Medicine Conference. 2016 December 7.

**Poster Presentations**

- [1] Epidemics 8, 8<sup>th</sup> international conference on Infectious Disease Dynamics, 2021
- [2] American Society of Tropical Medicine and Hygiene Annual Meeting, 2019
- [3] Epidemics 7, 7<sup>th</sup> international conference on Infectious Disease Dynamics, 2019

**Recognition, Fellowships & Awards**

- 1/2022 Dean's PhD Tuition Scholarship – Johns Hopkins University
- 9/2022 PhD Fellowship Awards for COVID Relief – Johns Hopkins University
- 4/2021 The Louis I. and Thomas D. Dublin Award – Johns Hopkins University
- 6/2017 US Embassy Award for creating maps and reports for visiting interagency team
- 4/2016 US Embassy Award for preparations concerning Dr. Tom Frieden's visit to Tanzania
- 5/2014 Wilbur G. Downs International Student Travel Fellowship – Yale University
- 5/2014 Coca-Cola World Fund Fellowship – Yale University
- 5/2013 Phi Beta Kappa Society – University of California, Berkeley
- 5/2013 College of Chemistry Senior Research Award – University of California, Berkeley

**Language**

Swahili	Speaking: <i>Beginner</i>	Reading: <i>Beginner</i>	Writing: <i>Beginner</i>
French	Speaking: <i>Beginner</i>	Reading: <i>Beginner</i>	Writing: <i>Beginner</i>

**Laboratory Experience**

- 3/2012 – Undergraduate Researcher, Francis Group, College of Chemistry,  
5/2013 University of California, Berkeley, Berkeley, CA
- 6/2011 – Undergraduate Researcher, Bissell Lab, Biological Systems and Engineering  
12/2011 Division, Lawrence Berkeley National Laboratory, Berkeley, CA
- 6/2010 – Undergraduate Researcher, Cochran Lab, Sackler School of Graduate Biomedical  
8/2010 Sciences, Tufts University, Boston, MA

**Teaching Experience**

- 6/2021 – Graduate Teaching Assistant, Epidemiology of Infectious Diseases  
7/2021 Graduate Summer Institute of Epidemiology and Biostatistics  
Johns Hopkins Bloomberg School of Public Health
- 1/2021 – Graduate Teaching Assistant, Concepts and Methods of Infectious Disease  
3/2021 Epidemiology  
Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
- 1/2020 – Graduate Teaching Assistant, Concepts and Methods of Infectious Disease  
3/2020 Epidemiology Department of Epidemiology, Johns Hopkins Bloomberg School of  
Public Health
- 10/2019 – Graduate Teaching Assistant, Epidemiologic Methods II  
12/2019 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
- 6/2019 Graduate Teaching Assistant, Longitudinal Data Analysis  
Graduate Summer Institute of Epidemiology and Biostatistics  
Johns Hopkins Bloomberg School of Public Health
- 1/2019 – Graduate Teaching Assistant, Epidemiologic Methods III  
3/2019 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health
- 10/2018 – Graduate Teaching Assistant, Epidemiology of Infectious Diseases  
12/2018 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

- 8/2014 – Graduate Teaching Fellow, Introductory Medical Statistics (STAT 105A)  
12/2014 Department of Statistics, Yale University, New Haven, CT
- 2/2014 – Personal Tutor, Advanced Placement Calculus and Chemistry  
5/2014 Ivy Labs, New Haven, CT
- 8/2010 – Study Group Lecturer and Drop-In Tutor, General Chemistry (CHEM 1A)  
5/2013 Student Learning Center, University of California, Berkeley, Berkeley, CA

**Volunteer / Extracurricular**

- Tutor at From Prison Cells to PhD (P2P, <https://www.fromprisoncellstophd.org/>)
- SOURCE HIV Testing and Counseling Volunteer
- Infectious Disease Epidemiology Journal Club Organizer
- Epidemiology Student Organization Social Chair

**Manuscript peer review**

*Annals of Internal Medicine, Clinical Infectious Diseases, and EBiomedicine*