RESEARCH ARTICLE

Predicting future community-level ocular Chlamydia trachomatis infection prevalence using serological, clinical, molecular, and geospatial data

Christine Tedijanto[1]*, Solomon Aragie[2], Zerihun Tadesse[2], Mahteme Haile[3], Taye Zeru[3], Scott D. Nash[4], Dionna M. Wittberg[1], Sarah Gwyn[5], Diana L. Martin[5], Hugh J. W. Sturrock[6], Thomas M. Lietman[1,7,8,9], Jeremy D. Keenan[1,7], Benjamin F. Arnold[1,7]

1 Francis I. Proctor Foundation, University of California, San Francisco, California, United States of America, **2** The Carter Center Ethiopia, Addis Ababa, Ethiopia, **3** Amhara Public Health Institute, Bahir Dar, Ethiopia, **4** The Carter Center, Atlanta, Georgia, United States of America, **5** Division of Parasitic Diseases and Malaria, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America, **6** Locational, Poole, United Kingdom, **7** Department of Ophthalmology, University of California, San Francisco, California, United States of America, **8** Department of Epidemiology and Biostatistics, University of California, San Francisco, California, United States of America, **9** Institute for Global Health Sciences, University of California, San Francisco, California, United States of America

* christine.tedijanto@ucsf.edu

OPEN ACCESS

Citation: Tedijanto C, Aragie S, Tadesse Z, Haile M, Zeru T, Nash SD, et al. (2022) Predicting future community-level ocular Chlamydia trachomatis infection prevalence using serological, clinical, molecular, and geospatial data. PLoS Negl Trop Dis 16(3): e0010273. https://doi.org/10.1371/journal.pntd.0010273

Editor: Ali M. Somily, King Saud University College of Medicine, SAUDI ARABIA

Received: October 10, 2021

Accepted: February 23, 2022

Published: March 11, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pntd.0010273

Data Availability Statement: Community latitude and longitude values have been modified to protect the privacy of study participants. The pre-specified

Abstract

Trachoma is an infectious disease characterized by repeated exposures to *Chlamydia trachomatis* (*Ct*) that may ultimately lead to blindness. Efficient identification of communities with high infection burden could help target more intensive control efforts. We hypothesized that IgG seroprevalence in combination with geospatial layers, machine learning, and model-based geostatistics would be able to accurately predict future community-level ocular *Ct* infections detected by PCR. We used measurements from 40 communities in the hyper-endemic Amhara region of Ethiopia to assess this hypothesis. Median *Ct* infection prevalence among children 0±5 years old increased from 6% at enrollment, in the context of recent mass drug administration (MDA), to 29% by month 36, following three years without MDA. At baseline, correlation between seroprevalence and *Ct* infection was stronger among children 0±5 years old (ρ = 0.77) than children 6±9 years old (ρ = 0.48), and stronger than the correlation between active trachoma and *Ct* infection (0-5y ρ = 0.56; 6-9y ρ = 0.40). Seroprevalence was the strongest concurrent predictor of infection prevalence at month 36 among children 0±5 years old (cross-validated R^2 = 0.75, 95% CI: 0.58±0.85), though predictive performance declined substantially with increasing temporal lag between predictor and outcome measurements. Geospatial variables, a spatial Gaussian process, and stacked ensemble machine learning did not meaningfully improve predictions. Serological markers among children 0±5 years old may be an objective tool for identifying communities with high levels of ocular *Ct* infections, but accurate, future prediction in the context of changing transmission remains an open challenge.

Author summary

Trachoma, one of the leading infectious causes of blindness globally, is targeted for

layers,

Each year, eight local nurses and other healthcare professionals were recruited to serve as trachoma graders m2000 System), which is highly sensitive and specific for *Ct* [22,23]. Groups of five samples, stratified by community and age group, were pooled for testing, and community-level *Ct* infection prevalence was estimated from pooled results using a maximum likelihood approach [24]. Swabs from positive pools were tested individually for 0±5-year-olds at all visits, for 6±9-year-olds at months 12, 24, and 36, and if >80% of pools for a cluster were positive for all other age groups and time points. Approximately 12% of samples from 6±9-year-olds with an equivocal or positive pooled result at baseline were also tested individually. Air swabs were collected in every cluster at the beginning and end of eachr variables were explored based on prior associations with trachoma or other infectious diseases (S1 Table). When possible, features were extracted and aggregated using Google Earth Engine [40], and means were used for spatial and temporal aggregation unless otherwise specified in S1 Table. All features were aggregated to a grid resolution of 2.5 arc minutes (approximately 4.5 km at the median latitude of the study area) based on the lowest resolution dataset (Terra-Climate) and reprojected to WGS84. Each community was assigned to the grid cell containing its household-weighted geographic centroid, defined as the median latitude and longitude across all households in the community.

Models were built using predictor variables measured over the same (ªconcurrentº) and prior (ªforward predictionsº) time periods. Time-varying features were summarized based on calendar year, with 2015 data considered ªconcurrentº with month 0 trachoma indicators and so on. Time-varying features were first aggregated by month and then summarized based on recency relative to the time of monitoring (e.g. last 1 month or December of the calendar year, last 2 months, up to 12 months). To reduce collinearity, we evaluated pairwise Pearson correlation coefficients between temporal summaries of the same variable and dropped the summary over fe3 0b [illegible] partitioned the study area into 12 15x15km blocks, each containing 1±8 spatially proximate communities. Communities in the same block were assigned to the same validation set, with some sets consisting of more than one block. This approach decreases spatial dependence between training and validation sets in the same fold and simulates prediction in a new, but geographically proximate, area. Predictive performance was assessed using cross-validated root-mean-square-error (RMSE) and R^2 [51], where R^2 was calculated as:

$$1 - \frac{\sum_{c} (p_{cm} - \hat{p}_{cm})^2}{\sum_{c} (p_{cm} - \bar{p}_{cm})^2}$$

95% confidence intervals for R^2 were estimated using the influence function [52,53]. Communities received equal weight in all validation metrics.

As this was a secondary analysis, the sample size was fixed at 40 communities per survey. To our knowledge, there are no methods available to estimate power for cross-validated error in prediction problems. Instead, we estimated the minimum detectable effect for the correlation analysis. Assuming a two-tailed alpha of 0.05, we had 80% power to detect a correlation of 0.43 or larger with 40 communities [54].

Results

Study population

Approximately thirty children from each of two age groups (0±5 years old and 6±9 years old) were randomly sampled from each community at baseline and follow-up and from

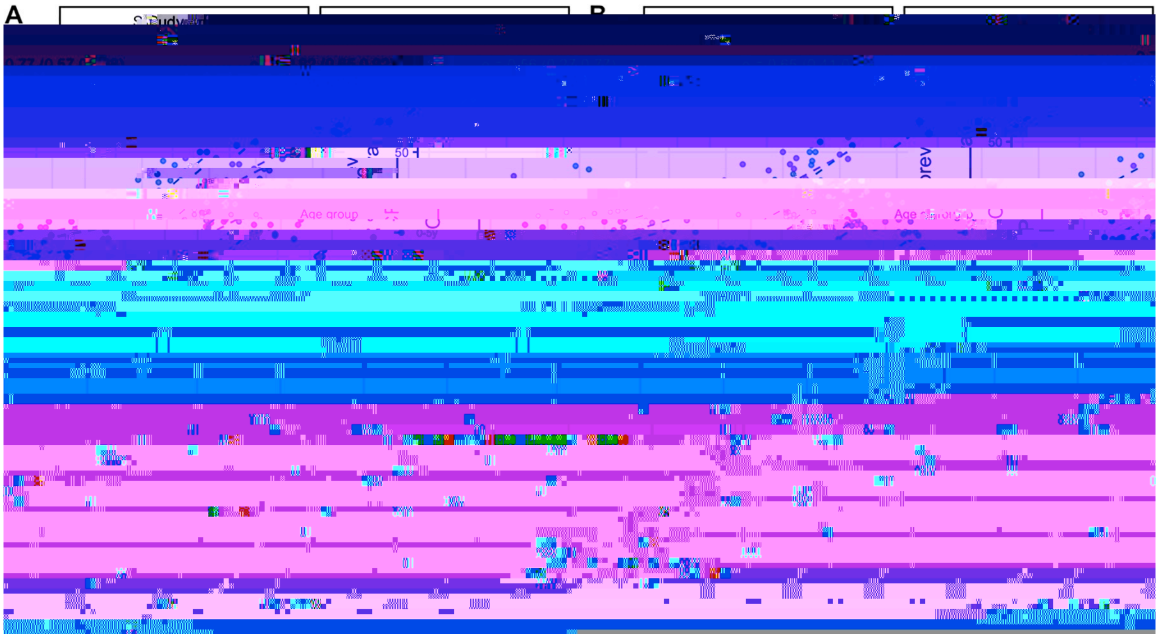

Fig 3. Correlations between trachoma indicators by age group and over time. Panels display Spearman rank correlations between community-level seroprevalence and PCR prevalence at study months 0 and 36 (A), active trachoma prevalence and PCR prevalence at months 0 and 36 (B), and PCR prevalence at month 36 and trachoma indicators measured at each survey across 40 study communities (C). Correlations are shown separately for 0±5-year-olds (green) and 6±9-year-olds (purple), and 95% confidence intervals were estimated from 1000 bootstrap samples. Serology data were not collected for a random sample of 6±9-year-olds at months 12 and 24.

https://doi.org/10.1371/journal.pntd.0010273.g003

correlations between trachoma indicators were more pronounced among younger children, potentially reflecting lower transmission in the presence of MDA and saturation in seroprevalence due to durable antibody responses among older children. Similar saturation dynamics may be at play for active trachoma, which has been shown to resolve slowly among children [55]. By month 36, when infections were higher across the study area (Table 1), correlations between trachoma indicators were similar across age groups (Fig 3A and 3B). Rank-preserving relationships between indicators at each time point and month 36 PCR prevalence were stronger for more proximate measurements, and this increase was more pronounced for PCR compared to active trachoma or serology (Fig 3C).

Concurrent and forward prediction of PCR prevalence

We predicted community-level infection prevalence using a range of model specifications and conducted spatial 10-fold cross-validation (CV) with 15x15 km blocks [49] to assess predictive performance using CV R^2 and root-mean-square-error (RMSE). Fig 4 presents results for models predicting PCR prevalence at month 36. ªConcurrentº predictions utilized trachoma indicators measured at month 36 and/or geospatial variables measured over the preceding year (2018), while ªforwardº predictions used covariates measured 12, 24, or 36 months in the past. Seroprevalence was the single strongest concurrent predictor of month 36 community-level PCR prevalence (CV R^2: 0.75, 95% confidence interval (CI): 0.58±0.85, CV RMSE: 0.10), substantially outperforming active trachoma prevalence (CV R^2: 0.37, 95% CI: 0.08±0.56, CV RMSE: 0.16) (Fig 4). When predicting 12 months into the future, all trachoma indicators performed moderately well, but predictive performance declined for longer time horizons across all model specifications. No model that we assessed had a CV R^2 significantly different from 0 (equivalent to an intercept-only or mean-only model) when predicting PCR prevalence 24 months or more into the future.

As anticipated by the weak spatial dependence in PCR prevalence (Fig 2), incorporation address variability in sample size, the number of Ct infections in each community was scaled to represent a sample of 30 individuals. At month 36, 80% of Ct infections were concentrated in just over half of the communities (23/40), and ordering communities by cross-validated concurrent predictions using seroprevalence identified infections more efficiently (i.e. in fewer communities, 25/40) than ordering them by predictions using visceral leishmaniasis reported 85.7% coverage of four-month-ahead 25±75% prediction intervals for case counts [60].

Our investigation b m 404e

S3 Table. Community-level seroprevalence across 40 study communities by antigen, age group, and study month.
(DOCX)

S1 Fig. Maps (A), variograms (B), and Moran's I (C) for seroprevalence among 0±5-year-olds at each study month. Maps display prevalence for 40 study communities at each follow-up visit, spatially interpolated over the convex hull using kriging. Variograms capture similarity between community-level prevalence measurements as a function of distance between community pairs (in km), with smaller semivariance values representing increased similarity. Exponential (magenta) and Matérn (green) models were fit to each empirical variogram, and the effective range (dashed vertical line) is defined as the distance at which the fitted model reaches 95% of the sill. The Monte Carlo envelope (gray shading) displays pointwise 95% coverage of 1000 permutations, representing a null distribution. Moran's I was calculated over 1000 permutations (gray bars, with observed value represented by red line), and a permutation-based p-value was calculated. The base map layer for panel A in this figure was downloaded from Stamen Maps (ªTerrainº) and is available under the CC BY 3.0 license.
(TIF)

S2 Fig. Maps (A), variograms (B), and Moran's I (C) for active trachoma prevalence among 0±5-year-olds at each study month. Maps display prevalence for 40 study communities at each follow-up visit, spatially interpolated over the convex hull using kriging. Variograms capture similarity between community-level prevalence measurements as a function of distance between community pairs (in km), with smaller semivariance values representing increased similarity. Exponential (magenta) and Matérn (green) models were fit to each empirical variogram, and the effective range (dashed vertical line) is defined as the distance at which the fitted model reaches 95% of the sill. The Monte Carlo envelope (gray shading) displays pointwise 95% coverage of 1000 permutations, representing a null distribution. Moran's I was calculated over 1000 permutations (gray bars, with observed value represented by red line), and a permutation-based p-value was calculated. The base map layer for panel A in this figure was downloaded from Stamen Maps (ªTerrainº) and is available under the CC BY 3.0 license.
(TIF)

S3 Fig. Correlations between PCR prevalence and antigen-specific seroprevalence by age group and over time. Panels display Spearman rank correlations between community-level Pgp3 seroprevalence and PCR prevalence at months 0 and 36 (A), CT694 seroprevalence and PCR prevalence at months 0 and 36 (B), and PCR prevalence at month 36 and seroprevalence measured at each follow-up visit across 40 study communities (C). Correlations are shown separately for 0±5-year-olds (green) and 6±9-year-olds (purple) when possible, and 95% confidence intervals were estimated from 1000 bootstrap samples. Serology data was not collected for a random sample of 6±9-year-olds at months 12 and 24.
(TIF)

S4 Fig. Spatio-temporal distribution of LASSO-selected geospatial predictor variables. Variables were estimated for 240 grid cells of 2.5 x 2.5 arc minutes (approximately 20 km^2 at the median latitude of the study area). Daily precipitation (A) and monthly night light radiance (B) averaged over the year were included in the final set of prediction models. The base map layer for this figure was downloaded from Stamen Maps (ªTerrainº) and is available under the CC BY 3.0 license.
(TIF)

S5 Fig. Cross-validated R^2 for models predicting community-level PCR prevalence among 0±5-year-olds at month 0 (A), at month 12 (B), at month 24 (C), at month 36 (D), and pooled across all months (E). Cross-validated R^2 (coefficient of determination), 95% influence-function-based confidence interval, and cross-validated root-mean-square error (RMSE, text label) are shown for each model specification. Blocks of size 15x15km were used for 10-fold spatial cross-validation. (D) is equivalent to Fig 4 in the main text and is included here for comparison.
(TIF)

S6 Fig. Cross-validated R^2 for stacked ensemble models predicting community-level PCR prevalence cross-validation. For predictions 36 months ahead, time could not be

trial. Lancet Glob Health. 2022 Jan 1; 10(1):e87±95. https://doi.org/10.1016/S2214-10

42. Breiman L. Stacked regressions. Mach Learn. 1996 Jul; 24(1):49±64.
43. Wolpert DH. Stacked generalization. Neural Netw. 1992 Jan 1; 5(2):241±59.
44. van der Laan MJ, Polley EC, Hubbard AE. Super Learner. 2007 [cited 2020 Nov 25]; Available from: https://biostat

64. Kim JS, Oldenburg CE, Cooley G, Amza A, Kadri B, Nassirou B, et al. Community-level chlamydial serology for assessing trachoma elimination in trachoma-endemic Niger. PLoS Negl Trop Dis [Internet]. 2019 Jan 28 [cited 2021 Mar 4]; 13(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6366708/ https://doi.org/10.1371/journal.pntd.0007127 PMID: 30689671

65. West SK, Munoz B, Mkocha H, Gaydos CA, Quinn TC. The effect of Mass Drug Administration for trachoma on antibodies to Chlamydia trachomatis pgp3 in children. Sci Rep. 2020 Sep 16; 10(1):15225. https://doi.org/10.1038/s41598-020-71833-x PMID: 32938957

66. Martin DL, Bid R, Sandi F, Goodhew EB, Massae PA, Lasway A, et al. Serology for Trachoma Surveillance after Cessation of Mass Drug Administration. Lietman TM, editor. PLoS Negl Trop Dis. 2015 Feb 25; 9(2):e0003555. https://doi.org/10.1371/journal.pntd.0003555 PMID: 25714363

67. West SK, Munoz B, Weaver J, Mrango Z, Dize L, Gaydos C, et al. Can We Use Antibodies to Chlamydia trachomatis as a Surveillance Tool for National Trachoma Control Programs? Results from a District Survey. Ngondi JM, editor. PLoS Negl Trop Dis. 2016 Jan 15; 10(1):e0004352. https://doi.org/10.1371/journal.pntd.0004352 PMID: 26771906

68. Migchelsen SJ, SepuÂlveda N, Martin DL, Cooley G, Gwyn S, Pickering H, et al. Serology reflects a decline in the prevalence of trachoma in two regions of The Gambia. Sci Rep. 2017 Nov 8; 7(1):15040. https://doi.org/10.1038/s41598-017-15056-7 PMID: 29118442

69. West SK, Zambrano AI, Sharma S, Mishra SK, Muñoz BE, Dize L, et al. Surveillance Surveys for Reemergent Trachoma in Formerly Endemic Districts in Nepal From 2 to 10 Years After Mass Drug Administration Cessation. JAMA Ophthalmol. 2017 Nov 1; 135(11):1141. https://doi.org/10.1001/jamaophthalmol.2017.3062 PMID: 28973295

70. Keenan JD, Lakew T, Alemayehu W, Melese M, Porco TC, Yi E, et al. Clinical Activity and Polymerase Chain Reaction Evidence of Chlamydial Infection after Repeated Mass Antibiotic Treatments for Trachoma. Am J Trop Med Hyg. 2010 Mar 1; 82(3):482±7. https://doi.org/10.4269/ajtmh.2010.09-0315 PMID: 20207878

71. Amza A, Kadri B, Nassirou B, Cotter SY, Stoller NE, West SK, et al. Community-level Association between Clinical Trachoma and Ocular Chlamydia Infection after MASS Azithromycin Distribution in a Mesoendemic Region of Niger. Ophthalmic Epidemiol. 2019 Jul 4; 26(4):231±7. https://doi.org/10.1080/09286586.2019.1597129 PMID: 30957594

72. Ramadhani AM, Derrick T, Macleod D, Holland MJ, Burton MJ. The Relationship between Active Trachoma and Ocular Chlamydia trachomatis Infection before and after Mass Antibiotic Treatment. PLoS Negl Trop Dis. 2016 Oct 26; 10(10):e0005080. https://doi.org/10.1371/journal.pntd.0005080 PMID: 27783678

73. Nash SD, Stewart AEP, Zerihun M, Sata E, Gessese D, Melak B, et al. Ocular Chlamydia trachomatis Infection Under the Surgery, Antibiotics, Facial Cleanliness, and Environmental Improvement Strategy in Amhara, Ethiopia, 2011±2015. Clin Infect Dis. 2018 Nov 28; 67(12):1840±6. https://doi.org/10.1093/cid/ciy377 PMID: 29741592

74. Odonkor M, Naufal F, Munoz B, Mkocha H, Kasubi M, Wolle M, et al. Serology, infection, and clinical trachoma as tools in prevalence surveys for re-emergence of trachoma in a formerly hyperendemic district. PLoS Negl Trop Dis. 2021 Apr 16; 15(4):e0009343. https://doi.org/10.1371/journal.pntd.0009343 PMID: 33861754

75. Clements ACA, Kur LW, Gatpan G, Ngondi JM, Emerson PM, Lado M, et al. Targeting Trachoma Control through Risk Mapping: The Example of Southern Sudan. PLoS Negl Trop Dis. 2010 Aug 17; 4(8):e799. https://doi.org/10.1371/journal.pntd.0000799 PMID: 20808910

76. Polack SR, Solomon AW, Alexander NDE, Massae PA, Safari S, Shao JF, et al. The household distribution of trachoma in a Tanzanian village: an application of GIS to the study of trachoma. Trans R Soc Trop Med Hyg. 2005 Mar 1; 99(3):218±25. https://doi.org/10.1016/j.trstmh.2004.06.010 PMID: 15653125

77. Diggle P, Lophaven S. Bayesian Geostatistical Design. Scand J Stat. 2006; 33(1):53±64.

78. SchéÂmann J-F, Sacko D, Malvy D, Momo G, Traore L, Bore O, et al. Risk factors for trachoma in Mali. Int J Epidemiol. 2002;(31):194±201. https://doi.org/10.1093/ije/31.1.194 PMID: 11914321

79. Bero B, Macleod C, Alemayehu W, Gadisa S, Abajobir A, Adamu Y, et al. Prevalence of and Risk Factors for Trachoma in Oromia Regional State of Ethiopia: Results of 79 Population-Based Prevalence Surveys Conducted with the Global Trachoma Mapping Project. Ophthalmic Epidemiol. 2016 Nov; 23(6):392±405. https://doi.org/10.1080/09286586.2016.1243717 PMID: 27820657

80. Hsieh Y-H, Bobo LD, Quinn TC, West SK. Risk Factors for Trachoma: 6-Year Follow-up of Children Aged 1 and 2 Years. Am J Epidemiol. 2000 Aug 1; 152(3):204±11. https://doi.org/10.1093/aje/152.3.204 PMID: 10933266

81. Phiri I, Manangazira P, Macleod CK, Mduluza T, Dhobbie T, Chaora SG, et al. The Burden of and Risk Factors for Trachoma in Selected Districts of Zimbabwe: Results of 16 Population-Based Prevalence

Surveys. Ophthalmic Epidemiol. 2018 Dec 28; 25(sup1):181±91. https://doi.org/10.1080/09286586.2017.1298823 PMID: 28532208

82. Alemayehu W,

100. Aybar C, Wu Q, Bautista L, Yali R, Barja A. rgee: An R package for interacting with Google Earth Engine. J Open Source Softw. 2020;

101. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010; 33(1):1±22. PMID: 20808728

102. Rousset F, Ferdy J-B. Testing environmental and genetic effects in the presence of spatial autocorrelation. Ecography. 2014; 37(8):781±90.

103. Coyle JR, Hejazi NS, Malenica I, Sofrygin O. sl3: Modern Pipelines for Machine Learning and Super