In disease control or elimination programs, diagnostics are essential for assessing the impact of interventions, refining treatment strategies, and minimizing the waste of scarce resources. Although high-performance tests are desirable, increased accuracy is frequently accompanied by a requirement for more elaborate infrastructure, which is often not feasible in the developing world. These challenges are pertinent to mapping, impact monitoring, and surveillance in trachoma elimination programs. To help inform rational design of diagnostics for trachoma elimination, we outline a nonparametric multilevel latent Markov modeling approach and apply it to 2 longitudinal cohort studies of trachoma-endemic communities in Tanzania (2000–2002) and The Gambia (2001–2002) to provide simultaneous inferences about the true population prevalence of *Chlamydia trachomatis* infection and disease and the sensitivity, specificity, and predictive values of 3 diagnostic tests for *C. trachomatis* infection. Estimates were obtained by using data collected before and after mass azithromycin administration. Such estimates are particularly important for trachoma because of the absence of a true "gold standard" diagnostic test for *C. trachomatis*. Estimated transition probabilities provide useful insights into key epidemiologic questions about the persistence of disease and the clearance of infection as well as the required frequency of surveillance in the postelimination setting.

### Abbreviations
- LMM, latent Markov model
- PCR, polymerase chain reaction
- TF, trachomatous inflammation, follicular
- TI, trachomatous inflammation, intense

An essential component of programs for the evaluation and improvement of global health is the availability of accurate diagnosis. Yet current diagnostics for many infectious diseases do not meet the needs of the developing world (1). Additionally, the lack of a diagnostic "gold standard" against which the accuracy of these tests would be evaluated compounds the problem. Trachoma, caused by ocular infection with the bacterium *Chlamydia trachomatis*, is the leading infectious cause of blindness in poor communities of developing countries (2) and is a disease with exactly these diagnostic challenges.

Although polymerase chain reaction (PCR)–based assays are considered to be very sensitive and specific for the diagnosis of *C. trachomatis* infection, they are currently unavailable in trachoma-endemic settings and too expensive to be used routinely as impact evaluation tools in large-scale treatment programs. Therefore, the methods currently used to evaluate the success of control strategies are clinical examinations to detect the presence of active disease (trachomatous inflammation, follicular (TF) and trachomatous inflammation, intense (TI)). The World Health Organization recommends annual mass drug administration with azithromycin...
wherever the district prevalence of TF in children aged 1–9 years is 10% or greater (3). However, the diagnostic performance of clinical examination is hindered by the frequent presence of active disease in the absence of infection and of infection in the absence of active disease (4–6). This is partly because clinical signs of trachoma may persist for many weeks after infection has been cleared (6–8) and also may result when infections other than Chlamydia cause trachoma-like inflammatory disease (9). Consequently, detection of active trachoma signs may not constitute the best tool to evaluate control interventions. Furthermore, policy makers require accurate surveillance of disease to decide when treatment aimed at eliminating infection can be discontinued and to detect reemergence in previously treated areas (10). Also of importance as implementation is rolled out is the question of whether and how the diagnostic performance of the available tests used at baseline may change with initial endemicity and after mass drug administration.

Latent variable modeling, that is statistical modeling that includes unobserved random variables that can alternatively be thought of as underlying parameters, has been used in medical research for the analysis of diagnostic accuracy and disease progression parameters in the absence of an acceptable gold standard (11, 12), as in the case of trachoma. For instance, Grassly et al. (13) used a multistate hidden Markov model to estimate the median duration of active disease and infection for different age groups by using a longitudinal data set followed up every 2 weeks over a 6-month period in The Gambia. Estimates were also obtained for the sensitivity and specificity of PCR or for clinical symptoms for infection, but without distinguishing between TF and TI. Hidden Markov models have been used to describe processes in which an individual moves through a series of states in continuous time, such as over the course of chronic and infectious diseases (14–18), most often by using a single diagnostic test. Through latent class analysis, 2 recent studies evaluated the performance of the trachoma diagnostics DNA-PCR, TF (i.e., detection of clinical symptoms of TF), and TI (i.e., detection of clinical symptoms of TI) (10) as well as DNA-PCR, RNA-PCR, and TF (19) by using cross-sectional data from pretreatment and posttreatment hyperendemic areas in Ethiopia, respectively. Latent class analysis allows the results of several diagnostic tests to be analyzed simultaneously, allowing for measurement error and using the information from all the diagnostic indicators to make a diagnosis for each individual in the study (20, 21).

In the current study, we developed latent Markov models (LMMs), which we applied to 2 longitudinal data sets from Tanzanian and Gambian communities with low baseline trachoma prevalence before and after mass administration of azithromycin. We then assessed the diagnostic accuracy of DNA-PCR, TF, and TI in the absence of a gold standard for baseline and follow-up time points. LMMs can be seen as an extension of latent class analysis for the analysis of longitudinal data, and they have been proposed for public health research in several studies (22–24). However, their application to the evaluation of diagnostic accuracy has not, to our knowledge, been previously investigated.

All models mentioned above—both hidden Markov models and LMMs (i.e., extensions of continuous and discrete Markov chains, respectively, that deal with measurement error) as well as latent class analysis—are mixture models. In the current study, we extended existing methodological research by fitting multilevel LMMs and by allowing some of the model parameters to differ across households. This approach is novel in that it takes into consideration the nested structure of the data and overcomes computational challenges associated with multilevel longitudinal mixture models; it allows for simultaneous evaluation of the 3 diagnostic tests with a test of measurement invariance as well as evaluation of the effect of a mass drug administration–based intervention on model-estimated infection and disease prevalence over time in the absence of a gold standard.

MATERIALS AND METHODS

Study areas, data collection, and analysis

Studies were conducted in East and West Africa (in the Kahe Mpya subvillage of the Rombo District, Tanzania and in the Upper Saloum District, The Gambia). The presence of ocular C. trachomatis infections at baseline and follow-up times was assessed through a DNA-PCR assay (Amplicor PCR; Roche Diagnostics Corp., Indianapolis, Indiana). Active disease was defined as the presence of TF and/or TI by using the World Health Organization’s simplified grading system (25). Detailed demographic information was collected, including individual’s age, sex, and household membership. Full descriptions of the study populations, structure of their communities, laboratory methods used, and summary statistics have been published elsewhere (7, 26–31). We fitted models to the principal reservoir of infection (27) (i.e., children aged <10 years) because trachoma control programs currently make mass drug administration decisions on the basis of the prevalence of active disease in children aged 1–9 years.

For all subjects, follow-up examinations and eye swabbing were performed by the same examiner at 2, 6, 12, 18, and 24 months after mass drug administration in Tanzania (during 2000–2002) and at 2, 6, 12, and 18 months in The Gambia (during 2001–2002). In Tanzania, only 1 person had a positive swab at 24 months (Figure 1); therefore, statistical analysis was restricted to the first 5 time points for this data set. A single round of mass administration of azithromycin was offered following baseline data collection to all residents of communities participating in these studies. Because the work presented here involves only further statistical analyses of previously published, deidentified data obtained in the original studies (28–31) (which had been granted ethical approval), additional ethical approval was not required.

Model specification for binary indicators

LMM consists of a structural model for the latent health states (analogous to the latent classes in latent class analysis) and a measurement model for the observed indicators (the 3
Latent Markov Model to Evaluate Trachoma Diagnostics 915

Figure 1. Descriptive (proportion positive) results according to the 3 diagnostic tests under investigation (DNA-PCR for Chlamydia trachomatis infection and TF and TI for signs of active trachoma disease) for children aged <10 years prior to and after mass azithromycin administration in A) Tanzania (2000–2002) and B) The Gambia (2001–2002). Numbers of children participating at each time point of the surveys are indicated in parentheses. PCR, polymerase chain reaction; TF, trachomatous inflammation, follicular; TI, trachomatous inflammation, intense.

diagnostic tools: PCR, TI, and TF) conditional on latent health state.

Let \( Y_{it} = (y_{it1}, \ldots, y_{itp}) \) be a response pattern for the \( i \)th individual at time \( t \) on \( P \) observed binary indicators (in this study the observed indicators are the 3 binary diagnostic tests, \( P = 3 \)) with values 0 and 1 indicating negative and positive diagnostic test results, respectively.

We assume that for each subject, the true underlying health state at each discrete time point \( t \) (where \( t = 1, \ldots, T \) and \( T = 5 \): baseline, 2, 6, 12, and 18 months) is explained by a latent categorical variable denoted by \( C \) with \( J \) latent health states.

The responses to the \( (P \times T) \) indicator variables are assumed to be independent conditional on the latent health state membership, which in our analysis implies that the results from the 3 diagnostics are assumed to be independent conditional on the true underlying health state both within and across time points. The latent categorical variable \( C_{it} \) depends on \( C_{i,t-1} \) but not on earlier latent categorical variables, known as the first-order Markov property. Under these assumptions, the probability of observing a particular response pattern \( Y \) for a randomly selected individual \( i \) is the following:

\[
P(Y_i = y_i) = \sum_{j_1=1}^J \cdots \sum_{j_T=1}^J P(C_{i,1} = j_1) \prod_{t=2}^T P(C_{i,t} = j_t|C_{i,t-1} = j_{t-1}) \prod_{p=1}^P \rho_{pt} |y_{ipt}^{(1-\gamma_{ipt})}(1 - \rho_{pt|jt})^{y_{ipt}}
\]

where

\[
P(C_{i,1} = j_1) = \frac{\exp(\gamma_{j_1})}{1 + \sum_{k_1=1}^J \exp(\gamma_{k_1})} = \eta_{j_1},
\]

represents the baseline health state prevalence at the first time point (i.e., baseline), and

\[
P(C_{i,t} = j_t|C_{i,t-1} = j_{t-1}) = \frac{\exp(\gamma_{j_t|j_{t-1}})}{1 + \sum_{j_{t-1}=1}^J \exp(\gamma_{j_t|j_{t-1}})}
= \tau_{C_{i,t-1}C_{i,t}}.
\]

represents the probability of a transition to latent health state \( j_t \) at time \( t \) conditional on membership in latent health state \( j_{t-1} \) at time \( t-1 \). \( \gamma_{j_t|j_{t-1}} \) stands for the coefficients from multinomial logistic regressions for transitions between \( t-1 \) and \( t \).

Furthermore, the probability of a negative response to \( p \)th indicator at time \( t \) conditional on latent health state \( j \) at time \( t \) is given by

\[
P(y_{ipt} = 0|C_{i,t} = j) = \frac{\exp(\delta_{ij})}{1 + \exp(\delta_{ij})} = \rho_{pt|jt}.
\]

The \( \rho \) quantities are known in the latent variable modeling literature as the item response probabilities, and they show precisely how the indicators (i.e., the diagnostic tests) measure the latent health states. In the case where the latent health states were dichotomous (identified as “not infected” and “infected”) and 0 represented negative diagnostic test results, equation 4 would represent the diagnostic specificity for infection when \( C_{i,t} \) is a “not infected” health state and \( P(Y_{ipt} = 1|C_{i,t} = j) \) would represent the sensitivity for infection of the 3 diagnostics, respectively, when \( C_{i,t} \) is an “infected” health state.

Depending on the number of latent health states (reflecting combinations of infected/not infected with diseased/not diseased), we derived estimates of sensitivities and specificities as well as positive and negative predictive values for all 3 diagnostic tools for ocular chlamydial infection on the basis of the estimates of \( \eta \) and \( \rho \) (equations 2 and 4) of relevant latent health states. For more technical details, see equations W6–W9 in the Web Appendix available at
Model selection

In mixture modeling, statistical theory determines the best model that combines goodness of fit and parsimony through minimum values for information criteria. In the current study, we explored these aspects among models that are biologically plausible for the epidemiological settings studied here (given the effect of mass drug administration on the natural history of trachoma). The inspection of information criteria informed our decisions (Web Table 1) and helped us explore the initial specifications of simple LMMs that did not account for household clustering of individuals. We made decisions with respect to the number of latent health states \( J \) required to obtain a good fit.

1. We then tested whether the transition probability matrix was stationary following treatment (i.e., 1 transition probability matrix between baseline and first follow-up and then another for subsequent transition periods vs. a separate transition probability matrix for each transition period). A global stationarity assumption (i.e., the equality of transition probability \( \tau \)'s across each of the transition points in LMMs) was not tested because this was not consistent with prior knowledge of mass drug administration impacts (28–30).

2. The hypothesis of measurement invariance of item response probabilities \( \rho \)'s across time was also tested. In general, if the \( \rho \)'s change over time, it is no longer as clear how to interpret the transition probabilities \( \tau \)'s because along with interpreting quantitative changes in latent health state membership, the meaning of the latent health states change over time.

Clustering of infection by household is a key characteristic of trachoma epidemiology (26). To allow for the nested structure of our 2 data sets (individuals within households), we expanded the selected model such that the categorical latent variables had random intercepts that varied between households. We used a discrete unspecified mixing distribution (i.e., nonparametric random-intercepts LMM) to examine if there was between-household variation in the prevalence of latent health states. We performed likelihood ratio tests comparing simple and multilevel LMMs. Further details of all models considered are given in the Web Appendix.

Maximum likelihood estimation and numerical challenges

Maximum likelihood estimation of mixture modeling is nontrivial and time consuming; it requires an iterative algorithm because mixture models generally exhibit multiple local maxima of the likelihood, and so fitting procedures
must be based on many starting values to ensure that a global maximum is found (32). In addition, maximum likelihood estimation of mixture models with random effects requires numerical integration, with each random effect with a non-0 level-2 variance contributing 1 dimension of integration (33, 34). Such problems have been circumvented for latent class analysis by the adoption of a nonparametric approach in which the computation burden is much less dependent on the number of random coefficients (35–37).

We adopted this approach for LMM, with which we were able to obtain a nonparametric random-intercepts LMM for only the first 4 of the 5 time points of the study (see Figure 3). We fitted all models by using MPlus, version 6.1, software (38), which automatically uses appropriate iterative algorithms and several sets of random starting values to reduce the probability of obtaining a local solution. The method of estimation was full information maximum likelihood, in which we assumed that missing data were missing at random. This means we assumed that missingness in these data was related to the conditions of the data collection rather than the latent health states of individuals. In Tanzania, there were very high follow-up levels for the first 5 years, whereas in The Gambia the main reason for people not being seen at a particular follow-up appointment was travel outside the study area. We also performed χ² tests for the missingness assumption. Such an approach allowed maximum use of data from individuals with incomplete (missing) data at the time points under study. Further technical details are given in the Web Appendix.

RESULTS

We fitted measurement-invariant LMMs (because the meaning of the latent health states also remains constant over time for these models) with 2 and 4 transition probability matrices for 2, 3, or 4 latent health states (Web Table 1). For Tanzania, our results indicated that a model with 2 transition probability matrices (1 for the 0- to 2-month follow-up and 1 for each subsequent follow-up period) and 3 latent health states was the most appropriate. We considered a similar model with 4 latent health states, but the item response probabilities did not point to meaningful interpretation of the latent health states; thus, we chose the model with 3 latent health states. For The Gambia, our results indicated that the model with 4 latent health states and 2 transition probability matrices was the most appropriate. The pattern of item response probabilities across the diagnostic tests (shown in Table 2) clearly differentiates among the latent health states and appears to correspond to different infection and disease states. For these selected models, there was no evidence that the missing at random assumption was violated (P values for χ² test for missing completely at random were 1.000). Nonparametric random effects on the latent health states were significant in both models (Web Table 2). The models failed to converge when data from the final observation were included in each country, which was most likely a result of the low prevalence of infection at this final follow-up time. We therefore present results based on the first 4 time points only. By examining the item response

![Figure 3](image-url)

Figure 3. Nonparametric multilevel latent Markov modeling path diagram with J latent health states (J = 4). The within-household model (level 1) is similar to the model in Figure 2 with 3 observed categorical indicators and 4 measurement time points; the 3 single filled circles at the bottom of each of the 4 Cₗ latent categorical variables represent the random means for the within-household latent health states (there are J − 1 random means). These random means are referred to as Cₗ # 1... Cₗ # 3 in the between-household model (level 2), and they vary across level 2 between households (cluster level) latent classes (labeled Cb in the above diagram). PCR, polymerase chain reaction; TF, trachomatous inflammation, follicular; TI, trachomatous inflammation, intense.

probabilities $\rho$ in Table 2, we assigned epidemiologically relevant interpretations to the latent health states. The first latent health state, labeled as “not infected and not diseased,” was characterized by very high probabilities of yielding negative results in both countries in all 3 examined diagnostic tests. Then, for Tanzania, “not infected and diseased” and “infected and diseased” can be assigned to the complementary corresponding probabilities of the second and third latent health states, respectively. In The Gambia, where a 4-state LMM was supported, the additional state appeared to correspond to “infected and not diseased.”

Having assigned interpretations to the latent health states, we examined the estimated prevalences $\eta$ of these latent health states over time (Figure 4). In both countries, the majority of children fell into the latent health state defined as “not infected and not diseased.” However, the proportion of


<table>
<thead>
<tr>
<th>Latent Health State</th>
<th>Tanzania</th>
<th>The Gambia</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I−/D−</td>
<td>0.990</td>
<td>0.982</td>
</tr>
<tr>
<td>0.972, 0.997</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>TF</td>
<td>0.999</td>
<td>0.982</td>
</tr>
<tr>
<td>0.937, 1.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>TI</td>
<td>0.999</td>
<td>0.982</td>
</tr>
<tr>
<td>0.937, 1.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>I+D−</td>
<td>0.172</td>
<td>0.172</td>
</tr>
<tr>
<td>0.056, 0.421</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>I−/D+</td>
<td>0.812</td>
<td>0.812</td>
</tr>
<tr>
<td>0.655, 0.988</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>I+D+</td>
<td>0.974</td>
<td>0.974</td>
</tr>
<tr>
<td>0.928, 0.999</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td>I−/D−</td>
<td>0.993</td>
<td>0.992</td>
</tr>
<tr>
<td>0.983, 1.000</td>
<td>0.992</td>
<td></td>
</tr>
<tr>
<td>I+D−</td>
<td>0.983</td>
<td>0.983</td>
</tr>
<tr>
<td>0.966, 0.999</td>
<td>0.992</td>
<td></td>
</tr>
<tr>
<td>I−/D+</td>
<td>0.983</td>
<td>0.983</td>
</tr>
<tr>
<td>0.966, 0.999</td>
<td>0.992</td>
<td></td>
</tr>
<tr>
<td>I+D+</td>
<td>0.983</td>
<td>0.983</td>
</tr>
<tr>
<td>0.966, 0.999</td>
<td>0.992</td>
<td></td>
</tr>
<tr>
<td>Abbreviations: CI, confidence interval; TF, trachomatous inflammation, follicular; TI, trachomatous inflammation, intense. a These parameters were estimated close to 1 or 0, so to avoid numerical instability in the estimation algorithm, MPlus software (Muthén &amp; Muthén, Los Angeles, California) fixed them automatically to 1 or 0, respectively. For this same reason, standard errors and confidence intervals are not provided.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Estimates of prevalences $\eta$ of latent health states over time as calculated from the nonparametric multilevel latent Markov model for A) Tanzania (2000–2002) and B) The Gambia (2001–2002). I−/D−, not infected and not diseased; I+D−, infected and not diseased; I−/D+, not infected and diseased; I+D+, infected and diseased.
those children who fell into this latent health state decreased in Tanzania after 6 months post–mass drug administration, whereas in The Gambia this proportion continued to increase over the study period. On the other hand, the proportion of children who fell into the latent health state, “not infected and diseased,” also increased in Tanzania after 6 months post–mass drug administration, whereas in The Gambia this proportion continued to decrease over the study period.

Furthermore, we examined the transitions between the latent health states over time by inspecting the elements of the transition probability matrices (Tables 3 and 4). Diagonal elements represent being in the same latent health state at time \( t \), conditional on being in that same latent health state at the previous time \( t_1 \). As expected, most of these elements were high in the nontreatment intervals. There was a higher probability of clearance of disease in uninfected individuals in The Gambia during both treatment and nontreatment intervals (\( \tau = 0.557 \) and 0.790, respectively) compared with Tanzania (\( \tau = 0.316 \) and 0.000, respectively).

Finally, by combining estimates of \( \eta \) and \( \rho \) and averaging them over the 4 time points of study here (equations W6–W9 in the Web Appendix), we were able to evaluate the accuracy of all 3 diagnostic tests over time for ocular chlamydial infection for Tanzania and The Gambia (Table 5).

## DISCUSSION

We applied nonparametric multilevel LMMs to 2 longitudinal data sets from Tanzanian and Gambian communities with low baseline trachoma prevalence before and after a round of mass azithromycin administration. This latent variable modeling approach yielded an average assessment of the diagnostic accuracy of PCR, TF, and TI for the detection of \( C. \) trachomatis infection in the absence of a gold standard.

We believe that our findings have important implications for monitoring and evaluation of trachoma control programs. For instance, TF is the measure on which mass drug administration decisions are often based.

The sensitivity and positive predictive value of clinical examination for infection were estimated to be very low in The Gambia, whereas the sensitivity of TI and positive predictive value of clinical examination (TF and TI) were very low in Tanzania, which is problematic; if treatment decisions are based solely on the simplified grading system, it could potentially lead to the decision to continue the annual mass administration of antichlamydial antibiotics for years after \( C. \) trachomatis has been eliminated from recipient communities (39). TF had higher sensitivity but lower positive predictive value and specificity for infection than did TI. The sensitivity for infection of DNA-PCR was estimated to be 77.8\% (95\% confidence interval, 41.5\%, 94.5\%) in Tanzania and 79.0\% (95\% confidence interval, 58.6\%, 99.4\%) in The Gambia, which is comparable with estimates in Ethiopia (19, 40). A model with variable sensitivity and specificity of the diagnostic tests (varying item response probabilities) over time was not supported by the data in The Gambia (Web Table 1), although this may be the result of low power given the low infection prevalence after treatment in this community. In Tanzania, there was support for variable performance of the diagnostic tests over time (Web Table 1).

### Table 3. Transition Probability Results \( \tau \) From the Nonparametric Multilevel Latent Markov Model for Tanzania, 2000–2002

<table>
<thead>
<tr>
<th>Time, ( t ) ( - ) 1</th>
<th>Treatment Interval (0–2 months), Time ( t )</th>
<th>Non-treatment Intervals (2–6 months and 6–12 months), Time ( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I^- / D^- )</td>
<td>1.000</td>
<td>0.768</td>
</tr>
<tr>
<td>( I^- / D^+ )</td>
<td>0.316</td>
<td>0.000</td>
</tr>
<tr>
<td>( I^+ / D^+ )</td>
<td>0.260</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Abbreviations: \( I^- / D^- \), not infected and not diseased; \( I^- / D^+ \), infected and not diseased; \( I^+ / D^- \), not infected and diseased; \( I^+ / D^+ \), infected and diseased.

### Table 4. Transition Probability Results \( \tau \) From the Nonparametric Multilevel Latent Markov Model for The Gambia, 2001–2002

<table>
<thead>
<tr>
<th>Time ( t ) ( - ) 1</th>
<th>Treatment Interval (0–2 months), Time ( t )</th>
<th>Non-treatment Intervals (2–6 months and 6–12 months), Time ( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I^- / D^- )</td>
<td>0.980</td>
<td>0.985</td>
</tr>
<tr>
<td>( I^- / D^+ )</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>( I^+ / D^+ )</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Abbreviations: \( I^- / D^- \), not infected and not diseased; \( I^- / D^+ \), infected and not diseased; \( I^+ / D^- \), not infected and diseased; \( I^+ / D^+ \), infected and diseased.

<table>
<thead>
<tr>
<th></th>
<th>Tanzania</th>
<th>The Gambia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PCR, %</td>
<td>95% CI</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>77.8</td>
<td>41.5, 94.5</td>
</tr>
<tr>
<td>Specificity</td>
<td>98.8</td>
<td>98.1, 99.5</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>98.3</td>
<td>96.3, 100</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>98.7</td>
<td>97.2, 100</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; PCR, polymerase chain reaction; TF, trachomatous inflammation, follicular; TI, trachomatous inflammation, intense.

δ The method has been used to approximate the standard errors for these functions of parameters as explained in the Web Appendix.

Because in the Tanzania analysis there is an I+/D+ latent health state but no I+/D− latent health state, the sensitivity estimates and their confidence limits are obtained directly from the item response probabilities and their confidence limits displayed in Table 2 by using the equation that sensitivity equals 1 minus the item response probability for I+/D+.

Because the normal approximation (i.e., estimate plus or minus 1.96 times the estimated standard error) does not take into account intrinsic constraints on parameter values, confidence intervals estimated in this way could extend beyond 0% or 100%. In these cases, we have used 0% as the minimum or 100% as the maximum.

In this case, the approximate standard error was so large that the normal approximation to the 95% confidence interval spanned the range from 0% to 100%. Clearly, there is little information in the data set about this positive predictive value.
thus, such interpretations should be made cautiously here. In addition, because of the designs of the longitudinal studies (specifically the sampling times of measurements performed), we recognize that these transition probabilities might well underestimate the true change between these latent health states over time (43). Because the time points are unevenly spaced, we could have fitted transition hazards rather than transition probabilities (in other words, hidden Markov models), but then only 1 diagnostic test would be evaluated because it is not computationally tractable to fit such models to multiple diagnostic tests simultaneously. Our fitted models have a separate transition probability matrix for the first time period (0–2 months, in which mass drug administration took place) and for subsequent transition periods. Thus, we have assumed the same transition probability in 4- and 6-month periods. We believe that this approximation is not crucial for our final inferences. However, in cases of less evenly spaced time points when simultaneous evaluation of multiple diagnostic tests with measurement error is not of primary interest, hidden Markov models might be more appropriate as a method of analysis. If there were more frequent evenly spaced measurements after mass drug administration, a more accurate picture of the posttreatment natural history of the trachoma might emerge. We propose that those designing longitudinal trachoma studies to inform control programs consider the measurement frequency carefully where resourcing could allow for more frequent follow-up.

In conclusion, this study has demonstrated that nonparametric multilevel LMMs can be used for the evaluation of diagnostic tools for trachoma in the absence of a diagnostic gold standard, even while allowing for the inherent hierarchical structure of these data: namely, repeated observations of children within households. We recommend that these methods be applied in settings with moderate and high baseline trachoma prevalence where longitudinal data can be generated before and after mass drug administration. Such research will aid trachoma program managers in their efforts to identify and characterize useful measures to evaluate control interventions. More generally, we recommend that LMMs be used for the evaluation of diagnostic tests of other diseases without gold standard diagnostic tools whenever longitudinal data are available, because such methodology has greater statistical power than does latent class analysis. Because LMMs use the additional information provided by repeated measurements, they might enable detection of additional latent health states or latent classes than if a latent class analysis was used at each time point. Furthermore, LMMs permit questions about changes in true health states and test the measurement invariance hypothesis of the diagnostic tests of interest over time, making them very useful tools for control program monitoring and evaluation research.

ACKNOWLEDGMENTS

Author affiliations: Medical Research Council Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Faculty of Medicine, Imperial College London, London, United Kingdom (Artemis Koukounari, Nicholas C. Grassly, Isobel M. Blake, María-Gloria Basáñez, Manoj Gambhir, Christl A. Donnelly); Department of Statistics, London School of Economics and Political Science, London, United Kingdom (Irini Moustaki); and Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom (David C. W. Mabey, Robin Bailey, Mathew J. Burton, Anthony W. Solomon).

This work was supported through a Medical Research Council Population Health Scientist Fellowship (A.K.). M. G.B. acknowledges funding from the Wellcome Trust and The Royal Society.

C.A.D. and N.C.G. thank the Medical Research Council for Centre funding. We thank Professor Bengt O. Muthén for help with the Mplus software and codes and Dr. Jouni Kuha for helpful advice with the Mplus codes and computational issues.

Conflict of interest: none declared.

REFERENCES


