Use of population health data to refine diagnostic decision-making for pertussis

Andrew M Fine,1 Ben Y Reis,1,2 Lise E Nigrovic,1 Donald A Goldmann,3 Tracy N LaPorte,4 Karen L Olson,1,2 Kenneth D Mandl1,2,5

ABSTRACT
Objective To improve identification of pertussis cases by developing a decision model that incorporates recent, local, population-level disease incidence.
Measurements Three models (based on clinical data only, local disease incidence only, and a combination of clinical data and local disease incidence) to predict pertussis positivity were created with demographic, historical, physical exam, and state-wide pertussis data. Models were compared using sensitivity, specificity, area under the receiver-operating characteristics (ROC) curve (AUC), and related metrics.
Results The model using only clinical data included cyanosis, cough for 1 week, and absence of fever, and was 89% sensitive (95% CI 79 to 99), 27% specific (95% CI 22 to 32) with an area under the ROC curve of 0.80. The model using only local incidence data performed best when the proportion positive of pertussis cultures in the region exceeded 10% in the 8–14 days prior to the infant’s associated visit, achieving 13% sensitivity, 53% specificity, and AUC 0.65. The combined model, built with patient-derived variables and local incidence data, included cyanosis, cough for 1 week, and the variable indicating that the proportion positive of pertussis cultures in the region exceeded 10% in the 8–14 days prior to the infant’s associated visit. This model was 100% sensitive (p<0.04, 95% CI 92 to 100), 38% specific (p<0.001, 95% CI 33 to 43), with AUC 0.82.
Conclusions Incorporating recent, local population-level disease incidence improved the ability of a decision model to correctly identify infants with pertussis. Our findings support fostering bidirectional exchange between public health and clinical practice, and validate a method for integrating large-scale public health datasets with rich clinical data to improve decision-making and public health.

INTRODUCTION
Bordetella pertussis outbreaks can infect hundreds of people across all age groups, though the infection is most dangerous for young infants.1–3 Pertussis is difficult to diagnose, especially in its early stages, and definitive test results are not available for several days. Timely administration of antibiotics decreases transmissibility of the disease.4 Most patients with cough do not have pertussis, but a missed case of the contagious disease is likely to have important consequences for the patient, her contacts, and the public health.5 A patient’s risk of exposure to infection varies by local disease burden,6 7 though clinicians rarely have ready access to information about epidemiologic context8—the recent regional incidence of an infectious disease—when making management decisions. The proliferation of real-time infectious disease surveillance systems9–10 and electronic laboratory reporting systems11 creates an opportunity to open a key communications channel between public health agencies and point-of-care providers. Currently there are no clinical decision-support systems that integrate public health incidence data into management algorithms in real time.12

Because of the temporal and geographic variability of pertussis outbreaks, the delay in diagnostic test results, and the personal and public health ramifications13–15 of incorrect management decisions at the point of care, pertussis is a prototypical disease for which real-time public health incidence data might inform, guide, and improve clinical decision-making. The purpose of this study is to quantify the value of recent, local disease incidence, derived from public health sources, in improving management of pertussis in the clinical setting.

METHODS
Design, setting, and subjects
A retrospective review was conducted of charts for infants tested for pertussis by culture, presenting to the pediatric emergency department (ED) of a large urban tertiary care US hospital from 1 January 2005 to 31 December 2007. The ED volume exceeds 50 000 patients per year. The study received institutional review board approval.

Inclusion and exclusion criteria
Subjects included all infants tested for pertussis by culture from 2005 to 2007. If a patient had multiple pertussis cultures from 2005 to 2007, only the first test was included.

Case definition
An infant was defined as pertussis-positive or pertussis-negative based on culture result, which is widely regarded as the gold standard.14 15 Alternate tests like PCR, serology, and direct fluorescent antibody (DFA) were not used in the case definition. Positive culture from a nasopharyngeal specimen is 100% specific for pertussis.4 16 Sensitivity, however, may be limited for several reasons including the organism’s fastidious nature, specimen collection technique, when the patient is tested in the course of the illness, and prior or concurrent use of antibiotics16 17 While PCR may have a better sensitivity, we did not rely on it because there is no FDA-approved test kit available, because test characteristics vary widely by laboratory and because
outbreaks have recently been attributed to PCR false positives.\textsuperscript{18} PCR may, in fact, be oversensitive, and requires correlation with at least 2 weeks of cough and paroxysm, whoop or post-tussive emesis,\textsuperscript{4} which are difficult to assess accurately in a retrospective review. Serology is not recommended for infants, and DFA is not widely available.\textsuperscript{19}

**Clinical data collection**

Demographics, signs, and symptoms commonly associated with infant pertussis, local disease incidence data and outcomes were collected for each patient.\textsuperscript{4, 20, 21} Demographics included visit date, gender, and age (months). Signs and symptoms included cough duration (days), fever duration (days), history of apnea, post-tussive emesis, cyanosis, seizure, and contact with a person with known pertussis. If the record did not contain information about these symptoms, they were coded as absent. Cough descriptors like paroxysm, staccato, and “whoop” were not included because they could not be measured accurately by chart review. Outcome data including antibiotic use, hospitalization, and mortality were collected to help describe the study population.

In the initial review, the pertussis culture result for each patient was obtained from the hospital laboratory information system. Subsequently, the chart abstractor (an attending physician specializing in pediatric emergency medicine) responsible for collecting and entering patient data into structured forms was blinded to the culture result. The culture result was accessible through a unique laboratory link to a PDF file from the external laboratory that performed the culture. These results were kept separated from the portion of the electronic chart used from the ED clinical encounter. No linkage between the culture result and the clinical portion of the chart was conducted until after all clinical charts had been reviewed. Historical and physical exam features were based on the EMR generated during the ED encounter. Outcome data were collected from the EMR, inpatient discharge summaries, and outpatient follow-up visits. To assess inter-rater reliability, a second abstractor (also an attending physician specializing in pediatric emergency medicine) reviewed a random sample of 7\% of charts.\textsuperscript{22} The two chart abstractors had over 90\% agreement (range 91–97\%) and K\textsuperscript{is} from 0.52 to 0.87 for all candidate predictors.

**Local disease-incidence data collection**

A query of the State Laboratory of the Massachusetts Department of Public Health database yielded 19,907 pertussis culture results from patients of all ages over the study period (2003–7). These data were obtained through a limited data sharing agreement. State data about cultures included date sent and culture result, but not demographics, clinical findings or outcomes.

Aggregate disease incidence variables were created for the number of pertussis cultures performed, the number of positives and the proportion positive at the state laboratory. Each of these variables was tabulated over a range of different timescales: 1–7, 8–14, 15–21, and 22–28 days prior to each visit date. Based on date of presentation, the corresponding public health incidence variables (number of cultures performed, positive, and proportion positive in the prior and cumulative 1–4 weeks) were assigned to each infant.

**Building the decision models**

The same sequence of steps was used to build three decision models: (1) “clinical only” model—candidate predictors included only clinical data based on demographics, history, and physical exam; (2) “local disease incidence” model—candidate predictors included only public health incidence data; and (3) “contextualized” model—all clinical and public health predictors were considered.

**Variable discretization and selection**

Dichotomous variables (history of apnea, post-tussive emesis, cyanosis, and seizure) associated with positive pertussis culture in the clinical data set were identified. Significance of association was tested with a chi² goodness-of-fit test (p<0.05). Continuous variables (duration of cough, duration of fever, and local disease incidence variables) were dichotomized at categorical cut-offs considered by the clinical investigators to be clinically useful and easy to remember (eg, cough at least 1 week, presence of fever, and proportion positive past 21 days >0.10).

In the multivariate analysis, candidate variables were entered into a forward stepwise logistic regression to identify independent predictors of infants testing pertussis positive. Cut-offs for entry and departure for the logistic regression model were 0.25 and 0.10.

For the local disease incidence model, each variable was considered for entry into the model as an independent predictor. Because of the interdependence of these variables, it was established a priori that no more than one candidate incidence variable would be contained in the final model. Thresholds were defined for the numbers of tests performed, positive, and proportion positive over 1–4 weeks. For proportion positive, thresholds were tested from 0.01 to 0.20 in increments of 0.01.

For the contextualized model, each clinical and local disease incidence variable was considered for entry into the multivariate model. Variables not included in the final clinical only or final local disease incidence model were still considered for inclusion into the contextualized model.

**Validation**

After selecting the best final model for each analysis (clinical, local disease incidence, or contextualized), a bootstrap validation was performed. Predictors that were selected in over 50\% of the 1000 bootstrap samples were retained in the final model.\textsuperscript{25–27}

**Measurement of model performance**

Sensitivity, specificity, positive and negative predictive values, area under the receiver-operating characteristics (ROC) curve (AUC), and percent correct classification were used to compare performance. The best model was defined as that with the greatest specificity among those with highest sensitivity, in order to minimize missed pertussis cases, and also minimize misclassification of those without pertussis.

**Comparing clinician performance with decision models**

Clinicians’ actual performance was compared with the clinical, local disease incidence, and contextualized models by measuring correct classification. Clinician performance of correct classification was judged by utilization or omission of antibiotics in the clinical encounter. The clinical actions taken, as determined by chart review, were compared with what would have been recommended based on the three decision models generated.

**RESULTS**

Four hundred and forty-three infants had a pertussis culture sent from 2003 to 2007, and 38 (8.4\%) tested positive. Nineteen thousand nine hundred and seven cultures were performed at the State Laboratory Institute of the Massachusetts Department of Public Health during the study; 1103 (5.5\%) tested positive. For these 19,907 cultures, the weekly proportion positive ranged
from 0 to 32% (mean 6.8%, median 5.6%, interquartile range (IQR) 3.3 to 8.8%). A mean of 4.2 cultures tested positive each week (median 4, range 0 to 20, IQR 2 to 6). Weekly and monthly (figure 1A,B) proportion positive at the state laboratory demonstrate that the timing, height, width, and total number of pertussis peaks vary annually.

Development of “clinical only” decision model

Univariate analysis

Infants testing positive for pertussis were younger, more likely to have a history of apnea or cyanosis, or cough for at least 1 week and were less likely to have fever than those who tested negative (table 1). There were no significant differences between those testing positive and negative for gender, history of post-tussive emesis or seizure, or exposure to a contact with known pertussis.

Multivariate analysis

History of cyanosis was the best predictor of pertussis, followed by history of cough for at least 1 week and absence of fever (table 2). Gender, history of post-tussive emesis or seizure, exposure to pertussis contact, and age in months were not included in the final clinical only model.

Figure 1  (A) Weekly proportion positive pertussis culture 2003–7. (B) Monthly proportion positive pertussis cultures, 2003–7. (A) and (B) demonstrate that the number, timing, height, and duration of pertussis peak vary annually. While the graph suggests some seasonality, it also shows that pertussis varies substantially from year to year.
Table 1  Subject characteristics by culture result

<table>
<thead>
<tr>
<th>Variable</th>
<th>Negative (n=405) N (%)</th>
<th>Positive (n=38) N (%)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male gender</td>
<td>216 (53%)</td>
<td>19 (50%)</td>
<td>0.40</td>
</tr>
<tr>
<td>Mean age (months)</td>
<td>3.4 (3, 1 to 5)</td>
<td>2.3 (2, 1 to 3)</td>
<td>0.02</td>
</tr>
<tr>
<td>Mean days of cough</td>
<td>8.8 (4, 1 to 10)</td>
<td>9.6 (7, 5 to 14)</td>
<td>0.71</td>
</tr>
<tr>
<td>Cough ≥1 week</td>
<td>147 (36%)</td>
<td>26 (68%)</td>
<td>0.002</td>
</tr>
<tr>
<td>History of apnea</td>
<td>72 (17%)</td>
<td>13 (34%)</td>
<td>0.04</td>
</tr>
<tr>
<td>Post-tussive emesis</td>
<td>85 (21%)</td>
<td>12 (32%)</td>
<td>0.22</td>
</tr>
<tr>
<td>History of cyanosis</td>
<td>73 (18%)</td>
<td>24 (63%)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>History of fever</td>
<td>86 (21%)</td>
<td>1 (2.6%)</td>
<td>0.019</td>
</tr>
<tr>
<td>History of seizure</td>
<td>4 (1.0%)</td>
<td>0 (0%)</td>
<td>0.38</td>
</tr>
<tr>
<td>Hospitalized</td>
<td>244 (59%)</td>
<td>31 (82%)</td>
<td>0.04</td>
</tr>
<tr>
<td>Died</td>
<td>6 (1.5%)</td>
<td>0 (0%)</td>
<td>0.88</td>
</tr>
<tr>
<td>Exposure to pertussis tests 2 weeks prior</td>
<td>12 (2.9%)</td>
<td>2 (5.3%)</td>
<td>0.42</td>
</tr>
<tr>
<td>Antibiotics</td>
<td>226 (55%)</td>
<td>33 (87%)</td>
<td>0.0009</td>
</tr>
<tr>
<td>Public health incidence mean percent positive</td>
<td>5.6 (4.8, 2.9 to 7.7)</td>
<td>9.9 (8.8, 5.8 to 14)</td>
<td></td>
</tr>
</tbody>
</table>

*Pertussis culture has 100% specificity and variable sensitivity (4, 16). t x², IQR, interquartile range.

Development of “local disease incidence” decision model

Selection of incidence variables

In the 1–4 weeks prior to each visit date, the ranges of mean numbers of cultures performed (102 to 110), positive (4.7 to 4.9) and proportion positive (0.058–0.060) showed a small variation. Because means, medians, ranges, and SD corresponded closely for 1–7, 8–14, 15–21, and 22–28 days, metrics from a single time interval were chosen to represent “local disease incidence.” Due to the time required to definitively process a pertussis culture,4 8–14 days (2 weeks) was chosen as the best metric available for a clinical application. A range of thresholds were examined to determine the cut-off that would optimize specificity for maximum sensitivity. The area under the ROC curve for these variables was maximized when the proportion positive from 8 to 14 days prior exceeded 0.10 (p<0.0001).

Development of contextualized decision model

Among clinical variables, presence of cyanosis and cough for at least 1 week met criteria for selection into the final logistic regression model (table 3). The proportion positive 8–14 days prior to the test date also met criteria for selection into the final logistic regression model. Proportion positive thresholds ranging from 0.01 to 0.20 in increments of 0.01 were tested. In conjunction with the clinical variables above, the maximal area under the ROC curve occurred with a cut-off of 0.10. The best contextualized model contained three variables—history of cyanosis, cough for at least 1 week, and proportion positive >0.10 eight to 14 days earlier. The incidence variable was a stronger predictor than any clinical factor considered, except for history of cyanosis.

Table 2  Multivariate logistic regression analysis (“clinical only” model)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>OR</th>
<th>95% CI</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyanosis</td>
<td>6.3</td>
<td>2.9 to 13.4</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Cough ≥1 week</td>
<td>3.1</td>
<td>1.4 to 7.0</td>
<td>0.0036</td>
</tr>
<tr>
<td>Absence of fever</td>
<td>6.5</td>
<td>1.3 to 118</td>
<td>0.019</td>
</tr>
</tbody>
</table>

Table 3  Multivariate logistic regression analysis with local disease incidence variables included (contextualized model)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>OR</th>
<th>95% CI</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyanosis</td>
<td>7.0</td>
<td>3.3 to 16</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Proportion positive prior 8–14 days &gt;10%</td>
<td>5.5</td>
<td>2.3 to 13</td>
<td>0.0001</td>
</tr>
<tr>
<td>Cough ≥1 week</td>
<td>3.3</td>
<td>1.5 to 7.9</td>
<td>0.0045</td>
</tr>
</tbody>
</table>

Validation

All predictors from the multivariate analyses were validated by the bootstrap method and retained in the final models. Cyanosis was selected in over 99%, public health pertussis proportion positive ≥0.10 eight to 14 days prior in over 90%, and cough for at least 1 week in over 80% of 1000 bootstrap samples.

Measurement of performance of decision models

The best model derived in the clinical only analysis (history of cyanosis, cough for at least 1 week and absence of fever), generated an area under the ROC of 0.80, with 89% sensitivity and 27% specificity (table 4). Addition of variables not significant in the univariate analysis (gender, history of post-tussive emesis, history of apnea, history of seizure, exposure to known pertussis case, and age under 3 months) did not improve the area under the ROC. The best local disease incidence model achieved only 13% sensitivity and 55% specificity. In the best contextualized model (history of cyanosis, proportion positive 8–14 days earlier ≥0.10, and cough for at least 1 week) the area under the ROC was 0.82 with 100% sensitivity and 38% specificity.

The contextualized model outperformed the clinical and local disease incidence models across all metrics (table 4). Compared with the clinical model, the contextualized model achieved superior sensitivity (99–100), specificity (27–58), PPV (12–18), NPV (96–100), and area under the ROC (0.80–0.82).

Comparing clinician performance with decision models

The percent of positives treated with antibiotics and percent of negatives not treated with antibiotics were compared with hypothetical outcomes generated by the three decision models (table 5). The contextualized model missed no patients with pertussis. Among models that did not miss any cases (100% sensitivity), the contextualized model classified the fewest patients (62%, 95% CI 57% to 67%) without pertussis, which would have resulted in the most judicious antibiotic use and correct categorization of infants. While clinicians did not do as well as the contextualized model, clinicians outperformed the clinical only model, misclassifying about the same number of patients with pertussis (11% vs 13%) but misclassifying fewer patients without the disease (60% vs 73%) (table 5).

Table 4  Performance of decision models with 95% CIs

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Clinical only</th>
<th>Local disease incidence only</th>
<th>Contextualized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>89 (79 to 99)</td>
<td>13</td>
<td>100 (92 to 100)</td>
</tr>
<tr>
<td>Specificity</td>
<td>27 (22 to 32)</td>
<td>53</td>
<td>38 (33 to 43)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>12 (8 to 16)</td>
<td>26</td>
<td>15 (11 to 19)</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>96 (92 to 99.96)</td>
<td>94</td>
<td>100 (98 to 100)</td>
</tr>
<tr>
<td>Area under receiver operator curve</td>
<td>0.80</td>
<td>0.65</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Sensitivity, specificity, positive and negative predictive value, and area under the receiver operating characteristics curve for the best model in each category. The contextualized model was superior to the clinical model for all metrics, and was statistically better for sensitivity (p<0.04), specificity (p<0.001), and negative predictive value (p<0.02).
It is becoming increasingly feasible to deliver the retrospective nature of the study also required an effort to use available high-quality data, we approached this problem by integrating a large statewide public health dataset with a detailed hospital-based clinical dataset to develop a decision model for a disease with major public health importance—pertussis.

**DISCUSSION**

Clinicians make critical decisions in the face of uncertainty, and typically rely on individual clinical experience and discussion with close colleagues when making decisions about diagnosis and treatment. Previously, we showed that local disease incidence information about meningitis from a single hospital provides valuable epidemiologic context and enhances a clinical decision model for distinguishing aseptic from bacterial meningitis. Here, we demonstrate for the first time how an external public health surveillance source improves a clinical decision model, by incorporating state-wide “epidemiologic context.” Previous prediction models, derived from small numbers of patients, have identified clinical predictors of infantile pertussis like cyanosis and cough, and some models have considered seasonality, but none have incorporated local disease incidence. Seasonality is not a substitute for accurate real-time information about pertussis incidence as pertussis outbreaks are sporadic and do not follow consistent seasonal or geographic patterns. In our analysis, epidemiologic context was stronger than all but one clinical predictor (cyanosis). This finding underscores the importance of “situational awareness” in the clinical setting. Understanding the epidemiologic context in which a patient presents may provide critical information about the etiology of the patient’s problem, but currently, this type of information is not formally processed, considered, or utilized in clinical decision-making.

Our findings support a general approach of estimating clinical risk of disease at the point of care, accounting for local disease incidence. This approach uses epidemiologic context in the clinical decision-making process rather than relying solely on history, physical exam, heuristics, and preliminary diagnostic test results. It is becoming increasingly feasible to deliver public health information to clinicians at the point of care. The emergence of robust, real-time surveillance systems, automated reporting to public health, and widespread adoption of electronic health records present opportunities for bidirectional communication between clinical practice and public health. Our study also promotes the value of disease reporting and surveillance at the state level.

We demonstrate a useful synergy between clinical and public health information in the generation and refinement of clinical decision rules. Public health data have not previously been used to generate decision models because, while they contain detailed information about those who test positive, they contain limited information about those testing negative. Public health efforts focus on tracking, interviewing, and following patients with reportable diseases, so public health data sets contain far greater detail about individuals who test positive. Data about those testing negative are limited even further by patient privacy laws, which prohibit collection of detailed information about people without the reportable disease. This unbalanced data stream creates a unique challenge to the use of public health datasets for the creation of decision models, which rely on rich information about patients both with and without the disease. In an effort to use available high-quality data, we approached this problem by basing the clinical models on patients who had pertussis tests ordered, potentially biasing toward subjects for whom clinicians already suspected of pertussis. While the contextualized model outperformed the clinical only model, the clinical only model is most limited by its retrospective nature. First, a prospectively derived clinical model where testing was based on symptoms and structured data were acquired systematically might improve the performance of a clinical model. Second, while the study was carried out at a single site, this site provides care for 75% of the children who live in and around this large metropolitan area. Third, the incidence data are state-wide, while the patients are from a single, large metropolitan area. Fourth, we relied on the most conservative method for evaluating pertussis—culture—because it is widely regarded as the gold standard. As delineated in the case definition section of the methods, PCR may be oversensitive, and requires correlation with at least 2 weeks of cough and paroxysm, whoop, or post-tussive emesis, which are difficult to assess accurately in a retrospective review. Serology is not recommended for infants, and DFA is not widely available. Fifth, the study was limited by a lack of immunization data on the subjects because primary care records were not accessible for these ED patients. Sixth, most patients did not have blood tests performed as part of the evaluation, and so lymphocytosis could not be included in the models; however, while lymphocytosis is classically associated with pertussis, it has been shown to be neither sensitive nor specific.

**CONCLUSION**

This study validates a scientific method for integrating incidence data into a clinical decision model and suggests that “epidemiologic context” could be an important component of future clinical decision-support systems. A software application integrated with an electronic health record might display data to physicians about ambient public health conditions and prompt appropriate management, treatment and reporting processes based on a calculation that considered patient factors in a specific epidemiologic situation. This important refinement of clinical decision-making requires communication between public health and clinical settings, and programs to enable integration of public health data within clinical environments.

**Funding** This work was supported by grants K01HG000055 and 1 P01 HK00088 from the Centers for Disease Control and Prevention and by G08LM009778 and R01 LM007677 from the National Library of Medicine.

**Competing interests** None.

**Ethics approval** The Committee on Clinical Investigation of Children’s Hospital Boston approved the study.

**Contributors** All authors made substantial contributions to conception, design, analysis, and interpretation of data. AMF and KOM drafted the manuscript, and all

---

**Table 5** Misclassification of best models versus actual clinician performance

<table>
<thead>
<tr>
<th>Model</th>
<th>Percent of patients pertussis positive but not treated (95% CI)</th>
<th>Percent of patients pertussis negative but treated (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinician’s actual treatment per electronic chart review</td>
<td>13% (2 to 23)</td>
<td>60% (55 to 65)</td>
</tr>
<tr>
<td>Best clinical only model</td>
<td>11% (1 to 21)</td>
<td>73% (69 to 78)</td>
</tr>
<tr>
<td>Best local disease incidence only model</td>
<td>61% (45 to 76)</td>
<td>15% (11 to 18)</td>
</tr>
<tr>
<td>Best contextualized model</td>
<td>0% (0 to 8)</td>
<td>62% (57 to 67)</td>
</tr>
</tbody>
</table>
authors were involved in revising it critically for important intellectual content and final approval. AMF is guarantor.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES