MAJOR ARTICLE

Quantifying Uncertainty: Physicians' Estimates of Infection in Critically Ill Neonates and Children

Joachim E. Fischer,[1] Stephan Harbarth,[2] Alexander G. Agthe,[4] Anne Benn,[1] Steven A. Ringer,[5] Donald A. Goldmann,[6] and Sergio Fanconi[3]

[1]Children's Hospital, University of Zurich, Zurich, [2]Geneva University Hospitals, Geneva, and [3]Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; [4]Johns Hopkins University Hospital, Baltimore, Maryland; [5]Brigham and Women's Hospital and [6]Children's Hospital, Harvard Medical School, Boston, Massachusetts

(See the editorial commentary by Mrus on pages 1391–3)

To determine the diagnostic accuracy of physicians' prior probability estimates of serious infection in critically ill neonates and children, we conducted a prospective cohort study in 2 intensive care units. Using available clinical, laboratory, and radiographic information, 27 physicians provided 2567 probability estimates for 347 patients (follow-up rate, 92%). The median probability estimate of infection increased from 0% (i.e., no antibiotic treatment or diagnostic work-up for sepsis), to 2% on the day preceding initiation of antibiotic therapy, to 20% at initiation of antibiotic treatment ($P < .001$). At initiation of treatment, predictions discriminated well between episodes subsequently classified as proven infection and episodes ultimately judged unlikely to be infection (area under the curve, 0.88). Physicians also showed a good ability to predict blood culture–positive sepsis (area under the curve, 0.77). Treatment and testing thresholds were derived from the provided predictions and treatment rates. Physicians' prognoses regarding the presence of serious infection were remarkably precise. Studies investigating the value of new tests for diagnosis of sepsis should establish that they add incremental value to physicians' judgment.

Decision-making in the presence of uncertainty is a core task of clinical medicine. A particularly difficult situation is the management of critically ill neonates and children who develop symptoms suggestive of serious infection. Because of the high risk associated with untreated bacterial infection, most clinicians favor early prescription of antibiotic treatment [1, 2]. Despite the resulting low threshold to initiate antibiotic treatment, some infants have conditions that escape early diagnosis. To improve the accuracy of diagnostic work-up, several new diagnostic tests have been suggested [3–8]. However, no single parameter has gained undisputed acceptance [9–12].

A potential reason for this failure to change clinical practice is the inherent incompatibility between real-life conditions and the artificial case-control design of most studies of diagnostic-test accuracy, in which the discriminative power of a new diagnostic test is derived from a subgroup of patients who satisfy unanimously accepted criteria for being clearly infected (cases) or not (controls) [13, 14]. This case-control design suffers from the potential overestimation of the performance of the diagnostic test being studied [15], since it omits ambiguous episodes of suspected infection and, more importantly, disregards any patient information that is present in addition to the diagnostic test under scrutiny. This stands in contrast to clinical practice, in which physicians derive clues about the presence or absence of infection from the patient's history, the clinical course, and laboratory findings. Moreover, the clinical

Received 8 October 2003; accepted 21 January 2004; electronically published 29 April 2004.

Financial support: the Bonizzi Theler Foundation and the Roche Research Foundation.

Reprints or correspondence: Dr. Joachim E. Fischer, Dept. of Neonatology and Pediatric Intensive Care, University Children's Hospital, University of Zurich, Steinwiesstr. 75, Zurich, Switzerland (fischer@ifv.gess.ethz.ch).

Clinical Infectious Diseases 2004;38:1383–90

1058-4838/2004/3810-0009$15.00

evaluation is constantly updated as new diagnostic information arises. Most of the previously published studies of diagnostic-test accuracy neither addressed this clinical evolution nor used a framework which allows integration of the available clinical judgement with the diagnostic test under scrutiny [14].

Hitherto, clinicians' prior probability estimates of the presence or absence of serious infection in critically ill infants and children have not been evaluated in a systematic fashion. We therefore conducted a prospective cohort study to obtain daily predictions and to evaluate the accuracy of physicians' estimates on the probability of serious infection, using the commonly applied case-control design of traditional studies of diagnostic-test accuracy. The secondary objective was to relate the obtained probability estimates to the observed clinical management decisions, in order to determine antibiotic-treatment and testing thresholds.

METHODS

Study design and population. In this prospective, observational cohort study, physicians estimated the prior probability of infection daily for every hospitalized patient. During a 3-month period, we enrolled all consecutive patients admitted to a tertiary neonatal intensive care unit with 28 level-III beds in Boston, Massachusetts, and to a tertiary pediatric intensive care unit with 19 level-III beds in Zurich, Switzerland. The 2 units provide the entire spectrum of neonatal and pediatric critical care, except for extracorporal membrane oxygenation.

Data collection. Three trained research fellows obtained estimates of the presence of bacterial infection from the clinicians responsible for the care of the patients daily (including weekends) after morning rounds (which included presentation of history, physical examination, and review of laboratory and radiographic data) and whenever antibiotic therapy was initiated for suspected infection. At the morning round after initiation of antibiotic therapy, physicians were asked to update their probability estimate, taking into account the new laboratory and clinical evidence obtained since initiation of therapy, but provided that blood culture results were still pending. For gauging these predictions against an external validity criterion, physicians also predicted the probability of bacterial growth in cultures of blood samples obtained during diagnostic work-up for sepsis. Of the participating physicians (12 attending critical-care physicians and 15 fellows), fellows with <1 year of intensive care experience (8 of 15 fellows) were required to consult their attending physician before providing a prediction. Only 1 daily prediction per patient was included in the analysis.

Because the best method to obtain probability estimates for rare events remains controversial [16], we pretested our case record form and offered 3 different types of scaling categories: estimation of the probability of infection expressed as odds, as percentage probability, or according to a visual analog scale [17]. The final, prevalidated scaling instrument asked for an exact probability expressed as a percentage, and it provided a number-entry field plus a row of check-boxes (with the values 0, 1, 2, 3, 5, 10, 15, 20, 25, 30, 40, 50, 75, and 100). The prediction instrument yielded a repeat-test reliability of 0.79 and an interrater reliability (i.e., rating of the same patient by 2 different physicians) of 0.73. Additional information collected for each patient included age, sex, weight, reason for hospital admission, antibiotic exposure, results of microbiologic cultures, and mode of ventilation. (The final version of the data collection form is available on request from J.E.F.)

Outcome assessment. After discontinuation of a patient's antibiotic therapy, 1 of the senior physicians who had not provided the prediction at initiation of antibiotic therapy and 1 of the investigators independently classified the outcomes related to each course of antibiotic treatment. Courses of antibiotic prophylaxis were excluded from this analysis. Adjudicators were blinded to the physicians' a priori predictions. Disagreement was resolved by consensus. Further details on the adjudication criteria have been reported elsewhere [1]. The outcome categories were as follows: "proven systemic bacterial infection," "proven localized infection," "probable infection," "viral infection" and "infection unlikely or absent." "Rule-out treatment" was defined as an episode of unlikely infection in which antibiotic therapy was discontinued within 48 h after being initiated and not restarted within 72 h.

Data analysis. Using the Kruskal-Wallis test, physicians' predictions were compared across different time-points during the patients' clinical course. General linear models were employed to explain the variance of these predictions. The logarithm of the prediction value served as the dependent variable, with clinical situation, age, sex, predicting physician, and study site as covariates. We considered the possibility that differences between the participating intensive care units and between physicians might affect predictions across clinical situations. Therefore, we tested 2 possible interaction terms: (1) interaction between the clinical situation and the participating study unit and (2) interaction between the clinical situation and the predicting physician. The final analysis presents the more parsimonious interaction term, which was interaction term (1), clinical situation and study unit.

We assessed the discriminative ability of physicians' predictions at the time of initiation of antibiotic therapy by means of logistic regression analysis [18], comparing predictions for patients later determined to have proven systemic bacterial infection with predictions obtained from patients with episodes classified as rule-out treatment. This approach provides comparability to similar studies of diagnostic-test accuracy in the field and is least prone to classification bias [3–7]. The analysis was repeated for predictions obtained during the day preceding

initiation of antibiotic therapy and for the updated predictions obtained during the morning round after initiation of antibiotic therapy. Sensitivity analyses used a broadened case definition, which included episodes of proven localized infection or probable infection. Using this broadened case definition, we checked the stability of the estimates after adjusting for potential confounding by participating unit, patient sex, and patient gestational age (modeled as a categorical variable). The Hosmer and Lemeshow goodness-of-fit statistics were employed to test the calibration of the models [19]. This method assesses whether a prediction model is well-calibrated across a possible range of predictions. A poor fit is indicated by a "significant" *P* value; a large *P* value indicates good calibration.

We derived thresholds from observed prescription and testing patterns by analyzing predictions provided before and at the time of initiation of antibiotic treatment. The proportion of patients receiving antibiotics was plotted against strata of predictions (percentage probabilities of 0, 1–3, 5–10, 10–20,

Table 1. Demographic and clinical characteristics of a cohort of pediatric patients in a study of the diagnostic accuracy of physicians' prognostications regarding serious bacterial infection.

Characteristic	Patients (n = 347)	Hospital days (n = 2785)
Male sex	195 (56.2)	1768 (63.5)
Patient age		
Premature infants		
Aged <28 weeks	38 (11.0)	733 (26.3)
Aged 28–31.9 weeks	31 (8.9)	246 (8.8)
Aged 32–36.9 weeks	43 (12.4)	188 (6.8)
Term newborns	93 (26.8)	394 (14.1)
Infants aged 1–12 months	53 (15.3)	746 (26.8)
Children aged 1–5 years	40 (11.5)	303 (10.9)
Children aged >5 years	49 (14.1)	175 (6.3)
Median age in months (interquartile range)	0.2 (0–13)	0.5 (0–5.4)
Main reasons for admission		
Prematurity	87 (25.1)	1018 (36.6)
Respiratory distress or respiratory disease	55 (15.8)	321 (11.5)
Other neonatal disorder	10 (2.9)	18 (0.6)
Congenital malformation, noncardiac	14 (4.0)	75 (2.7)
Congenital heart disease requiring surgery	32 (9.2)	428 (15.4)
Congenital heart disease without surgery	15 (4.3)	98 (3.5)
Surgery, noncardiac	33 (9.5)	158 (5.7)
Infectious condition	47 (13.6)	249 (8.9)
Neurological disorder	21 (6.1)	173 (6.2)
Impending circulatory failure	12 (3.5)	53 (1.9)
Trauma	10 (2.9)	34 (1.2)
Metabolic disorder	7 (2.0)	11 (0.4)
Other[a]	4 (1.1)	149 (5.4)
Ventilation		
Intratracheal mechanical	230 (66.3)	1604 (57.6)
Noninvasive	63 (18.2)	225 (8.1)
Antibiotic exposure		
Treatment for suspected infection[b]	188 (54.2)	1046 (37.6)
Perioperative and other prophylaxis[c]	47 (13.6)	170 (6.1)

NOTE. Data are no. (%) of patients or hospital days, unless indicated otherwise.

[a] One hundred forty-six hospital days were contributed by 1 male infant with Undine syndrome.

[b] Some patients received >1 course of treatment for suspected infection.

[c] "Other" includes 7 patients who received systemic prophylaxis because of severe immunosuppression.

Table 2. Daily predictions for different clinical situations in a cohort of pediatric patients in a study of the diagnostic accuracy of physicians' prognostications regarding serious bacterial infection.

Clinical situation	No. of predictions made	Predicted probability of infection, %			P[a]
		Median (IQR)	10th–90th percentile	Mean ± SEM	
No antibiotic therapy, no symptoms	1299	0 (0–2)	0–5	2.3 ± 0.2	Reference
No antibiotic therapy, no diagnostic testing	733	0 (0–1)	0–2	1.2 ± 0.2	<.01
24 h prior to initiation of antibiotic therapy	70	2 (0–10)	0–50	11.4 ± 2.6	<.01
At initiation of antibiotic therapy	162	20 (5–75)	2–100	38.0 ± 2.9	<.01
Stratified by time of initiation					
Initiation at admission	90	10 (3–75)	2–100	33.3 ± 4.1	Reference
Initiation during hospitalization	72	45 (5–75)	2–100	43.0 ± 5.2	<.01
Stratified by age at initiation					
Premature infants, aged <32 weeks	54	7 (4–30)	2–60	21.3 ± 3.4	Reference
Neonates, aged 32–44 weeks	34	4 (2–8)	0–10	8.0 ± 2.4	<.05
Infants, aged 1–12 months	52	75 (38–100)	5–100	62.7 ± 4.9	<.01
Children, aged >1 year	22	75 (25–100)	5–100	66.8 ± 8.0	<.01
Day after initiation of antibiotic therapy	181	10 (2–60)	0–100	31.0 ± 2.9	<.01
At discontinuation of antibiotic therapy	138	1 (0–2)	0–5	2.1 ± 0.3	>.2

NOTE. IQR, interquartile range.

[a] Group-wise comparisons with the reference category, using nonparametric tests.

20–40, 40–70, 70–100), and a regression curve was fit through these data points. We defined the treatment threshold as the predicted probability of infection that corresponded to a treatment rate of 50%. We derived the testing threshold from the results of all predictions obtained when patients did not receive antibiotics and did not have laboratory tests performed. The testing threshold was defined as the 75th percentile of these prediction values [14].

Standard definitions of sensitivity, specificity, and likelihood ratio were used [14, 18]. We constructed receiver operating characteristic (ROC) curves, which graphically represent the true-positive rate (sensitivity) and the false-positive rate (1-specificity) for a range of cutoff points [20]. The areas under the receiver operating characteristic curve (ROC_a) were determined using the algorithm suggested by Hanley and McNeil [21]. The greater the ROC_a value, the more discriminative the diagnostic test or prediction: an ROC_a value of 1.0 indicates a perfectly discriminative result, whereas an ROC_a value of 0.5 indicates a nondiscriminative result.

All analyses were performed using SAS software, version 6.12 (SAS Institute). Tests were 2-tailed, with $P < .05$ indicating statistical significance.

RESULTS

Three hundred forty-seven patients contributed 2785 hospital days to the study. Important demographic and clinical characteristics of the patient cohort are shown in table 1. Predictions were obtained for 2567 hospital days (follow-up rate, 92%). Data were missing and were omitted from the analysis for 218 hospital days because of early patient discharge ($n = 126$), periods of extreme workload ($n = 78$), or ambiguous marks on the data collection form ($n = 14$).

Outcomes. A total of 188 patients received 219 courses of antibiotic treatment for episodes of suspected infection, of which 183 courses were initiated after hospital admission. An additional 57 courses of antibiotics were prescribed as prophylaxis for 47 patients. Overall, patients were exposed to systemic antimicrobial agents during 1216 hospital days (44% of all hospital days). Of the 219 courses of antibiotic treatment, 30 (14%) were administered for episodes of ultimately proven systemic bacterial infection, 36 (16%) for episodes of proven localized infection, 29 (13%) for episodes of probable infection, 7 (3%) for episodes of viral infection and 117 (53%) for episodes of unlikely infection, of which 83 (71%) were for rule-out treatment episodes (38% of all courses).

Predictions. The predicted probability of infection increased as patients progressed from not being treated towards initiation of antibiotic therapy (table 2). The median probability estimate of infection increased from 0% (i.e., no antibiotic treatment or diagnostic work-up for sepsis), to 2% on the day preceding initiation of antibiotic therapy, to 20% at initiation of treatment ($P < .001$). Figure 1 presents predictions at different time-points (figure 1*A*) and stratified by outcome category (figure 1*B*). General linear models revealed that the total explained variance of the predicted probability of infection had a coefficient (R^2) of 0.22 ($P < .001$). The clinical situation (before, at, or after initiation of antibiotic therapy) was the single

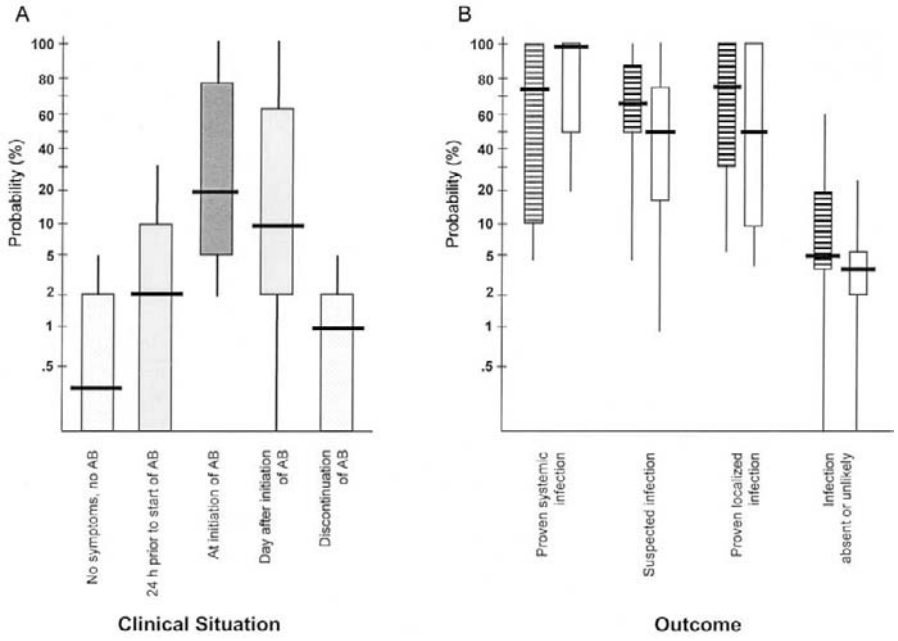

Figure 1. Physicians' predictions of the probability of serious bacterial infection in pediatric patients. *Horizontal black bars,* median values; *vertical boxes,* interquartile range; *vertical lines,* 10th–90th percentiles. *A,* Predictions at different time-points during the patients' clinical course. *B,* Predictions, stratified by outcome categories (see Methods for details). Box plots with horizontal shading show probability estimates at initiation of antibiotic therapy (AB); box plots without shading display the prognostications on the day after initiation of antibiotic therapy.

most important factor for estimating the probability of infection; it explained 15% of the observed variance. Other factors (i.e., study unit, age, and predicting physician) explained 7% of the variance. Addition of a term for interaction between the clinical situation and the study unit did not result in a significant effect.

Diagnostic accuracy. One day prior to initiation of antibiotic therapy, physicians were unable to discriminate between patients later determined to be clearly infected and those later classified as uninfected (ROC_a, 0.49). However, physicians' predictions on the day of initiation of antibiotic therapy discriminated well between episodes subsequently classified as proven systemic bacterial infection and those ultimately classified as rule-out treatment episodes (ROC_a, 0.88; 95% CI, 0.81–0.94; figure 2). With a cut-off point of the predicted probability of 25%, the sensitivity was 0.87 (95% CI, 0.65–0.97) with a corresponding likelihood ratio of 5.1 (95% CI, 3.5–7.9). A prediction of <25% yielded a specificity of 0.83 (95% CI, 0.73–0.90) and a likelihood ratio of 0.16 (95% CI, 0.1–0.2). Hosmer-Lemeshow χ^2 statistics revealed a good fit ($P = .63$). Results of sensitivity analysis with the broadened case definition (i.e., proven infection or probable infection) supported the stability of this estimate (figure 2). Adjusted analysis that controlled for potential confounding effects of age, sex, and study unit did not alter the estimate (ROC_a, 0.89; 95% CI, 0.82–0.94; $P = .91$, by Hosmer-Lemeshow χ^2 test). The discriminative ability of physicians' predictions made following the next morning round after initiation of antibiotic therapy increased to an ROC_a of 0.91 (95% CI, 0.84–0.96).

External validity criterion. Of 225 cultures of blood samples obtained during diagnostic work-up for sepsis, 23 grew bacteria, including 9 cultures of blood samples from 2 different puncture sites (a central line and a peripheral site) that yielded coagulase-negative staphylococci. Predictions were available for 205 (91%) of these cultures (blood culture positivity rate, 10.7%, 95% CI, 6.8–15.8%). The median predicted probability of a positive blood culture result was 15% (interquartile range, 10%–100%). Physicians showed a good ability to predict blood culture-positive sepsis, after adjusting for age and study unit (ROC_a, 0.77; 95% CI, 0.70–0.83; $P = .28$, Hosmer-Lemeshow χ^2 test).

Thresholds. Treatment and testing thresholds were derived from the observed antibiotic treatment rates in relation to different prediction strata. As shown in figure 3, treatment thresholds varied between patient subgroups, with lower thresholds for the more vulnerable patients. The median treatment threshold corresponded to a predicted probability of infection of 20% in newborns and of 50% in infants aged >1 month. The testing threshold was 1% in newborns and 2% in infants and children.

DISCUSSION

In this prospective study, physicians' a priori predictions at the time of initiation of antibiotic treatment discriminated well

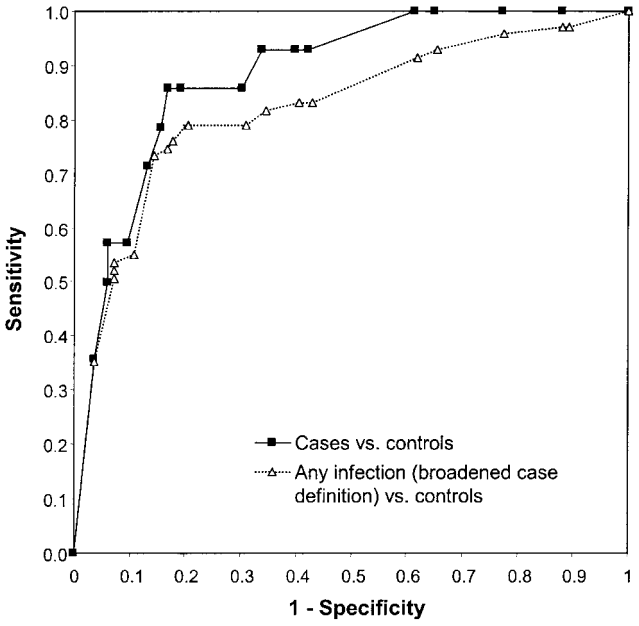

Figure 2. Receiver operating characteristic curves for the diagnostic accuracy of physicians' prognostications regarding serious bacterial infection in pediatric patients. The curve containing squares denotes the discrimination of cases with culture-proven systemic infection from control patients in whom infection was ultimately regarded as absent or unlikely. Triangles indicate the curve employing a broadened case definition comprising episodes ultimately classified as proven systemic bacterial infection, probable systemic infection, or proven localized infection.

between infected and noninfected patients. The diagnostic accuracy of physicians' judgment was within the range of published data for candidate biological markers of sepsis [3–7]. The estimate was stable in various sensitivity analyses. Moreover, physicians showed a good ability to predict positive results of blood cultures. A recent report on prediction of mortality in critically ill children showed similar agreement between the predictions made by experienced clinicians (ROC_a, 0.95) and the prognostication by the best available scoring system based on physiological data (ROC_a, 0.92) [22].

The key issue in the evaluation of candidate tests to use in the diagnostic work-up for infants or children suspected of having serious infection is not the diagnostic accuracy alone, but the diagnostic tests' clinical usefulness and added value. Approaches that combine clinical judgment based on all available information with the additional value of new diagnostic tests have been proposed, which use either Bayesian methods or logistic regression analysis [14, 18, 23]. The result of both approaches is an updated "posttest probability." If the posttest probability of serious infection exceeds a certain threshold, clinicians will initiate antibiotic treatment. A very low posttest probability may rule out infection sufficiently that physicians withhold antibiotics and refrain from further diagnostic testing [14]. Within this framework, a clinically useful test would move pretest probability estimates that are within the boundaries of uncertainty to posttest probability estimates that are outside those boundaries [14].

Current data on the clinical usefulness of new diagnostic markers of sepsis are limited. This is particularly important regarding the frequent situation when physicians suspect infection and, given current diagnostic strategies, initiate antibiotics. In the present study, for the majority of patients who were later determined to have proven infection, the predictions exceeded the observed treatment thresholds. This implies that positive results of laboratory tests would not have added information contributing to changed antibiotic-treatment decisions. However, negative results of powerful diagnostic tests could reassure physicians about deciding to withhold or to consider early discontinuation of antibiotic therapy [24]. Such tests require excellent sensitivity, to prevent errors in treatment of infected patients [25]. When combined with physicians' probability estimates, a test with a negative likelihood ratio of <0.1 would allow physicians to stop or withhold antibiotic treatment for 75% of all patients later determined to be uninfected. However, a test that would support withholding of antibiotic treatment when infection is already clinically suspected would have to satisfy performance characteristics exceeding those of any currently available biological marker of sepsis [14].

The traditional approach for evaluation of a new diagnostic test cannot assess its clinical usefulness, since it relies on the retrospective analysis of a subgroup of patients who satisfy unanimously accepted criteria for being clearly infected (cases) or not (controls) [13]. This study design inflates the estimates of a test's accuracy, since it does not evaluate the more difficult contrast among a group of similar patients with suspected infection [15]. For this reason, we suggest that future studies supplement the traditional approach by obtaining daily predictions on the probability of infection in a group of consecutive patients with clinically suspected infection. Combining physicians' probability estimates and test results will allow determination of posttest probabilities for every episode of suspected infection, including those with ambiguous classification.

A question of major importance remains: whether clinical vigilance and constant surveillance has the ability to advance the early diagnosis of infection in critically ill neonates and children [5], thus reducing the risk that asymptomatic infection will progress to septic shock. In our study, 1 day prior to initiation of antibiotic therapy, the predicted probability of infection for patients who subsequently had antibiotic therapy initiated was significantly higher than for patients who remained untreated. This indicates that physicians were probably aware of nonspecific clinical signs. However, at that time point, physicians' judgement alone was unable to identify patients who were subsequently classified as clearly infected.

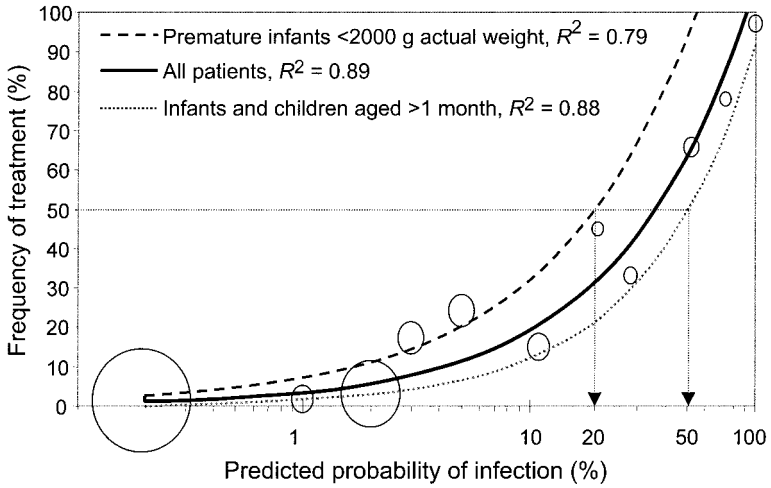

Figure 3. Treatment threshold curves and regression coefficients for diagnostic accuracy of physicians' predictions regarding the probability of serious bacterial infection in pediatric patients before and at initiation of antibiotic treatment. The horizontal axis denotes the predicted probability of infection, and the vertical axis indicates the related frequency of antibiotic treatment. The size of the circles corresponds to the number of predictions available from each stratum of predictions for calculation of the regression lines. *Bold regression line,* treatment rate–prediction curve for all patients; *left dotted line,* regression curve for premature infants; *right dotted gray line,* regression curve fitted through predictions obtained for infants and children aged >1 month; *thin vertical arrows,* 50 percent treatment rate, with the corresponding predicted probabilities.

The strengths of this study are that we developed treatment and testing thresholds from observed clinical data and that patient outcome was determined by investigators blinded to the probability estimates. However, some caveats of our study require consideration. First, treatment thresholds are a matter of subjective judgement and differ depending on the condition, the risk of therapy, the availability of tests, and the danger of the disease if left untreated. Therefore, the thresholds generated in this study may not be generalizable to adult patients. However, even if the results are not generalizable, the approach should be equally valid. Second, we cannot exclude a Hawthorne effect (i.e., observational bias) that possibly enhanced physicians' discriminative skills [26]. If such an effect was present, it did not translate into altered rates of prescription of antibiotics, as shown by a concurrently conducted observational study [1]. Finally, the limited number of patients with infection prevented us from controlling for the observed variance in predictions between physicians. However, the most likely consequence would be an underestimation of the accuracy of the predictions. Thus, we believe that our estimate of accuracy is conservative.

In summary, daily recording of clinicians' pretest probability estimates of infection is feasible and provides useful information for diagnostic-test accuracy studies. In the present study, physicians' prognostications regarding the presence of infection were remarkably precise. Studies investigating the value of new tests for diagnosing sepsis should establish that they add incremental value to physicians' judgment based on currently available clinical and laboratory parameters.

Acknowledgments

We are indebted to Dr. Francis E. Cook (Division of Clinical Epidemiology, Brigham and Women's Hospital, Boston, MA) for mentorship. We thank the physicians in both intensive care units for their dedicated participation and Michael Ramser (Zurich, Switzerland) for help in data collection.

References

1. Fischer JE, Ramser M, Fanconi S. Use of antibiotics in pediatric intensive care and potential savings. Intensive Care Med **2000**; 26:959–66.
2. Cordero L, Ayers LW. Duration of empiric antibiotics for suspected early-onset sepsis in extremely low birth weight infants. Infect Control Hosp Epidemiol **2003**; 24:662–6.
3. Berner R, Niemeyer CM, Leititis JU, et al. Plasma levels and gene expression of granulocyte colony-stimulating factor, tumor necrosis factor–α, interleukin (IL)–1β, IL-6, IL- 8, and soluble intercellular adhesion molecule–1 in neonatal early onset sepsis. Pediatr Res **1998**; 44:469–77.
4. Chiesa C, Panero A, Rossi N, et al. Reliability of procalcitonin concentrations for the diagnosis of sepsis in critically ill neonates. Clin Infect Dis **1998**; 26:664–72.
5. Kuster H, Weiss M, Willeitner AE, et al. Interleukin-1 receptor antagonist and interleukin-6 for early diagnosis of neonatal sepsis 2 days before clinical manifestation. Lancet **1998**; 352:1271–7.
6. Lacour AG, Gervaix A, Zamora SA, et al. Procalcitonin, IL-6, IL-8, IL-1 receptor antagonist and C-reactive protein as identificators of serious bacterial infections in children with fever without localising signs. Eur J Pediatr **2001**; 160:95–100.
7. Fischer JE, Benn A, Harbarth S, Nadal D, Fanconi S. Diagnostic accuracy of G-CSF, IL-8, and IL-1ra in critically ill children with suspected infection. Intensive Care Med **2002**; 28:1324–31.
8. Gonzalez BE, Mercado CK, Johnson L, Brodsky NL, Bhandari V. Early

markers of late-onset sepsis in premature neonates: clinical, hematological and cytokine profile. J Perinat Med **2003**; 31:60–8.
9. Fowlie PW, Schmidt B. Diagnostic tests for bacterial infection from birth to 90 days--a systematic review. Arch Dis Child Fetal Neonatal Ed **1998**; 78:F92–8.
10. Monneret G, Labaune JM, Isaac C, Bienvenu F, Putet G, Bienvenu J. Increased serum procalcitonin levels are not specific to sepsis in neonates. Clin Infect Dis **1998**; 27:1559–61.
11. Carlet J. Rapid diagnostic methods in the detection of sepsis. Infect Dis Clin North Am **1999**; 13:483–94.
12. Carroll AE, Silverstein M. C-reactive protein? Pediatrics **2002**; 110:422.
13. Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. J Clin Epidemiol **2003**; 56:1118–28.
14. Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. Intensive Care Med **2003**; 29:1043–51.
15. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA **1999**; 282:1061–6.
16. Bryant GD, Norman GR. Expressions of probability: words and numbers. N Engl J Med **1980**; 302:411.
17. Spector PE. Summated rating scale construction. Newbury Park: Sage Publications, **1992**.
18. Harbarth S, Holeckova K, Froidevaux C, et al. Diagnostic value of procalcitonin, IL-6, and IL-8 in critically ill patients admitted with suspected sepsis. Am J Respir Crit Care Med **2001**; 164:396–402.
19. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med **1997**; 16:965–80.
20. Michaud S, Suzuki S, Harbarth S. Effect of design-related bias in studies of diagnostic tests for ventilator-associated pneumonia. Am J Respir Crit Care Med **2002**; 166:1320–5.
21. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology **1982**; 143:29–36.
22. Marcin JP, Pollack MM, Patel KM, Ruttimann UE. Combining physician's subjective and physiology-based objective mortality risk predictions. Crit Care Med **2000**; 28:2984–90.
23. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med **1980**; 302:1109–17.
24. Escobar GJ, Zukin T, Usatin MS, et al. Early discontinuation of antibiotic treatment in newborns admitted to rule out sepsis: a decision rule. Pediatr Infect Dis J **1994**; 13:860–6.
25. Benitz WE, Han MY, Madan A, Ramachandra P. Serial serum C-reactive protein levels in the diagnosis of neonatal infection. Pediatrics **1998**; 102:E41.
26. De Amici D, Klersy C, Ramajoli F, Brustia L, Politi P. Impact of the Hawthorne effect in a longitudinal clinical study: the case of anesthesia. Control Clin Trials **2000**; 21:103–14.