Combining Clinical Risk Factors With Serum Bilirubin Levels to Predict Hyperbilirubinemia in Newborns

Thomas B. Newman, MD, MPH; Petra Liljestrand, PhD; Gabriel J. Escobar, MD

Objectives: (1) To validate a previously reported risk index for predicting total serum bilirubin (TSB) levels of 25 mg/dL (428 µmol/L) or higher; (2) to combine a subset of this index with TSB levels measured at less than 48 hours to predict subsequent TSB levels of 20 mg/dL (342 µmol/L) or higher.

Design: Nested case-control study using electronic and paper records (study 1). Retrospective cohort study using electronic records only (study 2).

Setting: Northern California Kaiser Permanente hospitals.

Patients: Subjects for both studies were newborns weighing 2000 g or more and of 36 weeks' or more gestation. The validation study included 67 cases born 1997-1998 who developed TSB levels of 25 mg/dL or higher at less than 30 days and 208 randomly selected control subjects. Subjects for study 2 were 5706 newborns who both were discharged from the hospital and had a TSB level measured at less than 48 hours.

Results: The risk index performed similarly in the validation group, born in 1997-1998, and the derivation group, born in 1995-1996 (area under the receiver operating characteristic curve = 0.83 vs 0.84). Of the 5706 newborns with TSB levels measured before 48 hours, 270 (4.7%) developed a TSB level of 20 mg/dL or higher. Of these, 254 (94%) had a TSB level at the 75th percentile or higher at less than 48 hours. The risk index improved prediction over the TSB level alone, largely owing to the effect of gestational age. For example, for those with a TSB level at the 95th percentile or higher at less than 48 hours, the risk increased from 9% for newborns born at 40 weeks' or more gestation to 42% for those born at 36 weeks.

Conclusion: Clinical risk factors significantly improve prediction of subsequent hyperbilirubinemia compared with early TSB levels alone, especially in those with early TSB levels above the 75th percentile.

Arch Pediatr Adolesc Med. 2005;159:113-119

With typical post-partum stays of 48 hours or less, outpatient follow-up is needed to identify the minority of infants in whom total serum bilirubin (TSB) levels will rise high enough to require treatment.1-4 The optimal timing and the importance of such follow-up visits vary depending on the infant’s risk of significant hyperbilirubinemia. Hence, several recent studies have looked at ways of predicting the risk of significant postdischarge hyperbilirubinemia by taking measurements before hospital discharge.

One approach, pioneered by Bhutani et al,5 involves measurement of a TSB level before hospital discharge and plotting the result according to the infant’s age in hours, to determine the percentile ranking for the infant’s TSB level. In the original study, this test performed exceptionally well—the area under receiver operating characteristic (ROC) curve to predict a TSB level of 17 mg/dL (291 µmol/L) or more was 0.95 (our calculation from the published results). (The area under the ROC curve, also called the “c-statistic” and abbreviated as “c,” is a measure of diagnostic discrimination, with 0.5 indicating discrimination no better than chance and 1.0 indicating perfect discrimination.6) However, differential follow-up may have falsely elevated this estimate.7 In a subsequent multicenter study using the same percentile graphs, the performance was not as good (c = 0.84 to predict a TSB level above the 95th percentile),8 but subjects whose TSB level already exceeded the 95th percentile at the first measurement were excluded from that study, which would reduce apparent prediction. Several other studies have shown similar results, with
To date, most studies of laboratory prediction of hyperbilirubinemia share 2 shortcomings. First, few consider information from the medical history and the findings from the physical examination that might independently predict subsequent risk. Second, the target for prediction, generally a TSB level above the 95th percentile, is not always clinically relevant, because many infants with TSB levels above the 95th percentile (about 17.5 mg/dL [299 µmol/L] for a baby 4 or more days old)\(^{3,15}\) do not require any treatment for jaundice.\(^{16}\)

Our group previously reported good performance (c=0.85) from a predictive index that used only items from the medical history and findings from the physical examination (exclusive breastfeeding, bruising, race, cephalhematoma, maternal age, sex, jaundice in previous sibling, and gestational age) to predict a TSB level of 25 mg/dL (428 µmol/L) or higher.\(^{17}\) That study had the advantage that the target TSB level was of definite clinical relevance. However, the index was not validated on a group of infants separate from the group on which it was derived, which may lead to overestimation of its discrimination.\(^{18}\) Moreover, in developing that predictive model we excluded infants with early jaundice or known high TSB levels. This was because of concern that early jaundice or early high TSB levels might lead to closer follow-up and/or treatment, which would attenuate the association between such levels and a TSB level of 25 mg/dL or higher. We subsequently found that many infants with early jaundice\(^{19}\) or known hyperbilirubinemia\(^{16}\) in the research population were not treated, diminishing this concern. The current study complements our previous work by testing the previously derived risk index on a new group of infants (ie, a separate validation data set), and by combining a subset of the risk index with hour-specific TSB percentile categories to predict TSB levels of 20 mg/dL or higher. As we have done previously,\(^{19}\) we used 2 different designs for the current study.

**STUDY 1: VALIDATION OF THE RISK INDEX FOR PREDICTING A TSB LEVEL OF 25 mg/dL OR MORE**

**Design**

For this study we required data available only from review of the medical records and, hence, used a nested case-control design.

The subjects were drawn from the cohort of infants born at Northern California Kaiser Permanente Medical Care Program hospitals from January 1, 1997, through December 31, 1998, who weighed at least 2000 g at birth, with a gestational age of at least 36 weeks (N=53997). The cases were infants with a maximum TSB level of 25 mg/dL or higher in the first 30 days after birth. We excluded 1 newborn because her maximum TSB level was listed in the computer as being before birth, and its exact timing could not be ascertained from the paper record. This left 67 cases. Controls (N=2088) were randomly sampled from the cohort. Cases and controls were identified and sampled in the same manner as the original derivation sample (born January 1, 1995, through December 31, 1996).\(^{17}\)

**Methods**

Methods for the 2 studies reported herein will be described separately. The institutional review boards for the protection of human subjects at the Kaiser Permanente Medical Care Program and the University of California, San Francisco, approved the abstraction of paper and electronic records used for both studies.

**Variables**

We used data electronically available on the entire birth cohorts to compare those born in 1995-1996 to those born in 1997-1998, as previously described and validated.\(^{15,17,20}\) Medical records analysts abstracted data on gestational age, breastfeeding, bruising, and cephalhematoma from paper medical records, as previously described.\(^{17}\) In contrast to our earlier work, we did not use family history of jaundice as a predictor because we discovered instances in which well-intentioned medical records analysts had entered this information based on rehospitalization rather than birth hospitalization data. Because family history data would more likely have been elicited and recorded for newborns rehospitalized for jaundice, which included almost all cases and few controls, including this variable could lead to bias. Except for this change, we calculated the risk index exactly as in the previous study, and defined and excluded newborns with early jaundice in the same way.

**Statistical Analysis**

We compared characteristics of the derivation and validation cohorts using \(\chi^2\) and rank sum tests and quantified the prediction of the modified risk index using the area under the ROC curve.

**STUDY 2: COMBINING A SUBSET OF THE RISK INDEX WITH TSB LEVELS TO PREDICT RISK OF TSB 20 mg/dL OR MORE**

**Design**

To investigate the effect of combining the electronically available subset of the risk index with TSB levels obtained at less than 48 hours, we used a retrospective cohort design.

**Subjects**

Subjects were drawn from the cohort of infants born at Northern California Kaiser Permanente hospitals from 1995 through 1998 weighing at least 2000 g at birth, with a gestational age of at least 36 weeks (N=105 384). We restricted attention to the 5711 newborns discharged at less than 48 hours after birth who had a TSB level measured before 48 hours. Because only 14 newborns in this group went on to develop a TSB level of 25 mg/dL or higher, we lowered the target for prediction to a TSB level of 20 mg/dL or higher. Five of the 5711 newborns already had TSB levels of 20 mg/dL or more at less than 48 hours. These were excluded, leaving 5706 subjects for the study, of whom 270 (4.7%) developed a TSB level of 20 mg/dL or higher at age 48 hours or older.
Variables

We obtained TSB levels, dates, and times on the 5706 qualifying subjects from electronic databases, as previously described. To classify TSB levels into age-specific percentile groups, we used data from a graph of TSB percentiles by hour of age from Bhutani et al. The lines for the percentiles on that graph are approximately linear for the first 48 hours. We visually estimated that the slopes for the 95th, 75th, and 40th percentile lines were 0.21, 0.17, and 0.14 mg/dL per hour, and that the intercepts were 3.1, 2.5, and 2.0 mg/dL, respectively.

The procedure above allowed categorization of TSB levels into 4 percentile groups (<40th, 40th-74th, 75th-94th, and ≥95th). This lead to some loss of information, because within each group, TSB percentiles varied. In particular, the group with TSB levels at or above the 95th percentile included some newborns close to the 95th percentile, and others well above the 99th percentile. To avoid this loss of information, we also transformed TSB levels into age-specific z scores by assuming TSB levels up to the 95th percentile were approximately normally distributed. Under this assumption, the 40th percentile was about 0.253 SD below the median, and the 95th percentile was about 1.645 SDs above the median. Hence, the hour-specific standard deviation was approximated by the difference between the 95th and 40th percentiles, divided by (0.253+1.645), and the median was just the 40th percentile +0.253 times that standard deviation. An hour-specific z score for the TSB level could thus be computed by subtracting the observed value from the calculated median for that age and dividing by the calculated standard deviation. If more than 1 TSB level was obtained before 48 hours, only the first one was used for analyses.

To determine the extent to which prediction from TSB percentiles might be enhanced by clinical information, we obtained data on the mother’s age and self-reported race, the newborn’s gestational age, and diagnosis of “scalp injury at birth” (International Classification of Diseases, 9th Revision, Clinical Modification; code 767.1). For the 5706 subjects described above. We used these to create a partial risk index based on data available electronically. This partial risk index was categorized and cross-tabulated with the hour-specific TSB percentile, we chose categories of the risk index that most closely approximated those of the TSB percentiles—that is, the 40th, 75th, and 95th percentiles of the risk index defined the 4 groups.

Statistical Analysis

To obtain the ROC curve for simultaneous use of the partial risk index and early TSB level, we used both as predictors in a logistic regression analysis—the partial risk index itself (range, –8 to 16) and the TSB percentile group as a single variable coded from 1 to 4. Using 3 indicator variables for the TSB percentile group did not improve prediction. We used Stata 8 software for all analyses. We used the ROCCOMP command to compare areas under ROC curves. This command considers the non-independence of curves measured on the same set of subjects; hence, P values for comparing 2 areas may be much smaller than might be expected from looking at their confidence intervals.

Table 1. Modified Risk Index for Predicting Hyperbilirubinemia in Infants Who Do Not Have Early Jaundice*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Points</th>
<th>Available Electronically?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclusive breastfeeding at hospital discharge</td>
<td>6</td>
<td>No</td>
</tr>
<tr>
<td>Bruising noted</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>Asian race</td>
<td>4</td>
<td>Yes</td>
</tr>
<tr>
<td>Cephalhematoma noted</td>
<td>3</td>
<td>ICD-9-CM code 767.1 used</td>
</tr>
<tr>
<td>Mother aged ≥25 y</td>
<td>3</td>
<td>Yes</td>
</tr>
<tr>
<td>Male sex</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>Black race</td>
<td>–2</td>
<td>Yes</td>
</tr>
<tr>
<td>Gestational age</td>
<td>2 × (40 − GA)</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Abbreviations: GA, gestational age; ICD-9-CM, International Classification of Diseases, 9th Revision, Clinical Modification.

*The modified risk index was calculated as the sum of all characteristics that apply to the patient, except that points for gestational age were assigned based on twice the difference from 40 weeks. The score is identical to that previously reported except that the item for jaundice in a previous sibling was deleted. For the partial risk index used in the retrospective cohort study, breastfeeding and bruising were also deleted and a diagnosis of scalp injury at birth (ICD-9-CM code 767.1) was substituted for cephalhematoma.

RESULTS

STUDY 1: VALIDATION OF PREVIOUSLY REPORTED RISK INDEX

The 1997-1998 study cohort was similar to the 1995-1996 cohort for serum bilirubin testing and serum bilirubin levels (Table 2). In both periods about 0.14% of the newborns developed a TSB level of 25 mg/dL or higher and about 2% developed a TSB level of 20 mg/dL or higher, the vast majority before 7 days of age. The mean length of hospital stay was about 7 hours longer in 1997-1998, but the proportion of newborns with procedure codes for inpatient phototherapy was similar in the 2 periods.

The performance of the modified risk index (not including a family history of jaundice) for prediction of the 67 patients with a TSB level of 25 mg/dL or higher born in 1997-1998 was similar to that for the 73 patients born in 1995-1996 from which the risk index was derived; the areas under the ROC curves were 0.83 (95% confidence interval [CI], 0.77-0.89) and 0.84 (95% CI, 0.79-0.89), respectively (P = .08 for difference in areas; Figure 1). Eliminating the family history variable had little effect on prediction of a TSB level of 25 mg/dL or higher because it was coded as present in only 10% of the cases. (The area under the ROC curve for the 1995-1996 cases declined from the previous reported value of 0.85 to 0.84).

STUDY 2: COMBINING THE CLINICAL RISK INDEX WITH EARLY TSB LEVELS TO PREDICT RISK OF A TSB OF 20 mg/dL OR MORE

For newborns with TSB measurements and hospital discharge at less than 48 hours (N = 5706), the risk of a documented postdischarge TSB level of 20 mg/dL or more ranged from 0.5% for an early TSB level below the 40th
percentile, to 13.8% for values at the 95th percentile or above (Table 3). Of the 270 newborns who developed a TSB level of 20 mg/dL or higher, the early TSB level for 254 (94%) was at the 75th percentile or higher, compared with 2950 (55%) of 5436 whose TSB level remained lower than 20 mg/dL.

The area under the ROC curve for the TSB percentile category alone was 0.79 (95% CI, 0.77-0.81). Combining the partial risk index with early TSB levels enhanced prediction (Figure 2). Within each percentile category, there was a 5- to 15-fold increase in risk of developing a TSB level of 20 mg/dL or higher for those with a risk index of 10 or higher compared with those with a risk index below 4. However, the absolute risk remained low (<2.5%) for newborns whose early TSB level was below the 75th percentile. In contrast, if the initial TSB level was at the 75th percentile or higher, large absolute differences by clinical risk index were apparent. For example, the proportion reaching a TSB level of 20 mg/dL or more for those with an early TSB level at the 95th percentile or higher was as high as 36% if the risk index was 10 or more, and as low as 6% if the risk index was less than 4. This last value (6%) was lower than that for an early TSB level at the 75th to 94th percentile with a risk index of 7 or more. Similar results are apparent using gestational age alone, rather than the risk index. The risk of developing a TSB level of 20 mg/dL or higher for a 36-week newborn with an early TSB level at the 75th to 94th percentile was greater than that for a full-term newborn with a TSB level at the 95th percentile or higher (Figure 3).

Considering those with early TSB levels above the 95th percentile as a single group wastes information, because some in that group have much higher TSB levels than others. Thus, prediction was enhanced by using hour-

### Table 2. Comparison of Original Derivation (1995-1996) and Validation (1997-1998) Cohorts

<table>
<thead>
<tr>
<th>Variable</th>
<th>1995-1996 Cohort, No. (%)</th>
<th>1997-1998 Cohort, No. (%)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of newborns in cohort</td>
<td>51 387</td>
<td>53 997</td>
<td>NA</td>
</tr>
<tr>
<td>At least 1 TSB level sent</td>
<td>13 181 (25.7)</td>
<td>14 053 (26.0)</td>
<td>.16</td>
</tr>
<tr>
<td>Highest recorded TSB level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\geq$15 mg/dL</td>
<td>4780 (9.3)</td>
<td>5455 (10.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>$\geq$20 mg/dL</td>
<td>1002 (1.9)</td>
<td>1098 (2.0)</td>
<td>.33</td>
</tr>
<tr>
<td>$\geq$25 mg/dL</td>
<td>75* (0.15)</td>
<td>68* (0.13)</td>
<td>.38</td>
</tr>
<tr>
<td>Age $\geq$7 d for highest TSB level, if the TSB level is $\geq$20 mg/dL</td>
<td>68/1002 (6.8)</td>
<td>75/1098 (6.8)</td>
<td>.97</td>
</tr>
<tr>
<td>Average No. of TSB tests per patient</td>
<td>0.64</td>
<td>0.66</td>
<td>.05†</td>
</tr>
<tr>
<td>Average length of hospital stay, h</td>
<td>38.8</td>
<td>45.7</td>
<td>&lt;.001†</td>
</tr>
<tr>
<td>Treatment with (inpatient) phototherapy</td>
<td>1430 (2.8)</td>
<td>1455 (2.7)</td>
<td>.38</td>
</tr>
</tbody>
</table>

Abbreviations: NA, not applicable; TSB, total serum bilirubin.

SI conversion: To convert TSB to micromoles per liter, multiply by 17.1.

Two of the 75 infants in the 1995-1996 cohort were excluded because paper records revealed that their gestational ages or birth weights did not meet inclusion criteria. One of the 68 infants in the 1997-1998 cohort was excluded because the exact timing of the maximum TSB level could not be ascertained.†Rank sum test.
We found that performance of a modified risk index to predict a TSB level of 25 mg/dL or higher derived from a nested case-control study of newborns born in 1995-1996 performed similarly on infants from the same hospitals born in 1997-1998, with an area under the ROC curve of 0.83 to 0.84. We also found that even the limited set of clinical risk factors available electronically, particularly gestational age, significantly improved prediction over the hour-specific TSB percentile. The main effect was on newborns whose initial TSB level was at the 75th percentile or higher—the risk was low among those whose initial TSB level was below the 75th percentile, even when risk factors were present.

A second limitation is the reliance on retrospective data. This was necessary because, although jaundice is common, TSB levels of 20 to 25 mg/dL or higher are rare, making a prospective study infeasible. For the retrospective cohort study, we were restricted to data available in the Northern California Kaiser Permanente databases. The partial risk index thus did not include data on breastfeeding—one of the strongest predictors of hyperbilirubinemia in our previous study.

Neither the original nor the partial risk index included results of blood types and direct antiglobulin tests or the severity of jaundice on physical examination, as these were not consistently obtained or recorded. In addition, measurement error for data items like cephalhematoma and gestational age might be worse for data collected retrospectively rather than prospectively. All of these omissions and errors would attenuate the apparent predictive power of a clinical risk index based upon retrospective data.

Third, the extent to which both risk factors and early TSB levels predict subsequent hyperbilirubinemia is also likely to be underestimated if high TSB levels and/or high levels of risk factors frequently led to interventions that prevented or attenuated the severity of hyperbilirubinemia. The higher the target TSB level for prediction, the more likely this is to be a problem, because very high TSB levels result not only from biologic predisposition to hyperbilirubinemia, but also from health services factors such as lack of early monitoring and treatment.

Thus, this bias would be greater for prediction of a TSB level of 25 mg/dL or higher than prediction of a TSB level of 20 mg/dL or higher. The approximately 40% risk of developing a TSB level of 20 mg/dL or higher in the highest risk groups (Figure 2 and Figure 3) and the relatively low levels of use of phototherapy at moderate TSB levels in this organization during the study period also argue against this problem.

Many studies have addressed the use of early serum bilirubin levels to predict subsequent hyperbilirubinemia. Results across studies are not strictly comparable because of different inclusion criteria, different target TSB levels for prediction at different ages, and varia-
ations in timing and classification of predischarge TSB results. Nevertheless, one consistent finding, confirmed in the current study, is that the negative predictive value of a TSB level below the 40th to 50th percentile for age is high—that is, newborns with these levels are at low risk of subsequent hyperbilirubinemia (however defined) [5,8,13,23,24]. On the other hand, because of the low prevalence of hyperbilirubinemia and the lack of specificity of the tests, unless the TSB is very high, the positive predictive value of TSB levels above this range tends to be low. In the current study, the most striking effect of the clinical risk data was increasing the positive predictive value of a TSB level above the 95th percentile.

Although many previous studies have examined clinical risk factors for hyperbilirubinemia, few have combined clinical and laboratory information. Seidman et al [29] studied clinical risk factors combined not only with early TSB levels, but also with the change in TSB levels over the first 2 days. However, the logistic model presented, which includes odds ratios for the day 1 TSB level, the change in TSB from day 1 to day 2, and a dichotomous variable for day 1 TSB levels exceeding 5 mg/dL, produced results that are difficult to interpret. Sarici et al [11] examined both TSB levels and clinical risk factors, but not both simultaneously. In that study, only the history of a previous sibling with jaundice was a significant predictor of subsequent hyperbilirubinemia, perhaps because the group studied was newborns with an ABO blood group incompatibility with their mother. Stevenson et al [8] took a different approach: they sought to determine whether an additional laboratory test, the ETCOc, added to the TSB level for prediction of hyperbilirubinemia. They found that addition of the ETCOc did not improve prediction, for reasons well summarized by Valaes. [14] This suggests clinical risk factors (primarily gestational age in the current study, but presumably also breastfeeding) may contribute more to decreased excretion of serum bilirubin than to increased production, as increased production would have been captured by ETCOc.

Our results have important implications for the current controversy regarding universal predischarge serum bilirubin level screening. While confirming that TSB measurements done before 48 hours are strong predictors of subsequent TSB levels of 20 mg/dL or higher, our results also suggest the need for caution. Use of predischarge TSB tests in isolation will lead to suboptimal decisions about future testing and follow-up. For example, additional TSB levels might erroneously be perceived as necessary in newborns who are at low risk, like those with TSB level in the 75th to 94th percentile but with a risk index of less than 4. Suboptimal decisions could also occur if, for example, an initial TSB level below the 40th percentile led clinicians or parents to ignore clinical risk factors (or subsequent significant jaundice) under the assumption that a significant risk of hyperbilirubinemia had been ruled out. These results highlight the need to base treatment and follow-up decisions not just on laboratory results, but on the medical history and the findings from the physical examination as well.

Accepted for Publication: August 4, 2004.
Correspondence: Thomas B. Newman, MD, MPH, Department of Epidemiology and Biostatistics, University of California, San Francisco, Box 0560, San Francisco, CA 94143 (newman@itsa.ucsf.edu).
Previous Presentation: Presented in part at the Pediatric Academic Societies Meeting, Baltimore, Md; May 6, 2002; and Pediatric Academic Societies Meeting, Seattle, Wash; May 5, 2003.
Funding/Support: This work was supported by grant M01 RR01271, Pediatric Clinical Research Center, from the National Institutes of Health, Bethesda, Md; grant R01 NS39683 from the National Institute of Neurologic Diseases and Stroke, Bethesda; and a grant from the David and Lucille Packard Foundation, Los Altos, Calif.
Acknowledgment: We thank M. Jeffrey Maisels, MB, BCh, and anonymous reviewers for suggestions on the manuscript.

REFERENCES

19. Newman TB, Liljestrand P, Escobar GJ. Jaundice noted in the first 24 hours af-

Announcement

Online Submission and Peer Review System Available. The Archives of Pediatrics & Adolescent Medicine editorial office has introduced an online manuscript submission and peer review system developed by eJournalPress that will serve the needs of authors, reviewers, and editors. The new system went live on February 7. See http://archpediatrics.org for more detailed information.