

Systematic Review of Screening for Bilirubin Encephalopathy in Neonates

abstract

CONTEXT: Severe neonatal hyperbilirubinemia is associated with chronic bilirubin encephalopathy (kernicterus).

OBJECTIVE: To systematically review the effectiveness of specific screening modalities to prevent neonatal bilirubin encephalopathy.

METHODS: We identified studies through Medline searches, perusing reference lists and by consulting with US Preventive Services Task Force lead experts. We included English-language publications evaluating the effects of screening for bilirubin encephalopathy using early total serum bilirubin (TSB), transcutaneous bilirubin (TcB) measurements, or risk scores. Severe hyperbilirubinemia was used as a surrogate for possible chronic bilirubin encephalopathy, because no studies directly evaluated the latter as an outcome. We calculated the sensitivity and specificity of early TSB, TcB measurements, or risk scores in detecting hyperbilirubinemia.

RESULTS: Ten publications (11 studies) were eligible. Seven (2 prospective) studies evaluated the ability of risk factors ($n = 3$), early TSB ($n = 3$), TcB ($n = 2$), or combinations of risk factors and early TSB ($n = 1$) to predict hyperbilirubinemia (typically TSB > 95th hour-specific percentile 24 hours to 30 days postpartum). Screening had good ability to detect hyperbilirubinemia: reported area-under-the-curve values ranged between 0.69 and 0.84, and reported sensitivities and specificities suggested similar diagnostic ability. Indirect evidence from 3 descriptive uncontrolled studies suggests favorable associations between initiation of screening and decrease in hyperbilirubinemia rates, and rates of treatment or readmissions for hyperbilirubinemia compared with the baseline of no screening. No study assessed harms of screening.

CONCLUSIONS: Effects of screening on the rates of bilirubin encephalopathy are unknown. Although screening can predict hyperbilirubinemia, there is no robust evidence to suggest that screening is associated with favorable clinical outcomes. *Pediatrics* 2009;124:1162–1171

AUTHORS: Thomas A. Trikalinos, MD, Mei Chung, MPH, Joseph Lau, MD, and Stanley Ip, MD

Tufts Evidence-Based Practice Center and Center for Clinical Evidence Synthesis, Institute for Clinical Research and Health Policy Studies, Boston, Massachusetts

KEY WORDS

hyperbilirubinemia, kernicterus, bilirubin encephalopathy, screening

ABBREVIATIONS

Tufts EPC—Tufts Evidence-Based Practice Center
AHRQ—Agency for Healthcare Research and Quality
USPSTF—US Preventive Services Task Force
TcB—transcutaneous bilirubin
TSB—total serum bilirubin
AUC—area under the curve
LR—likelihood ratio

The findings and conclusions are those of the authors and do not necessarily represent the views of the AHRQ. No statement in this article should be construed as an official position of the AHRQ or US Department of Health and Human Services.

www.pediatrics.org/cgi/doi/10.1542/peds.2008-3545

doi:10.1542/peds.2008-3545

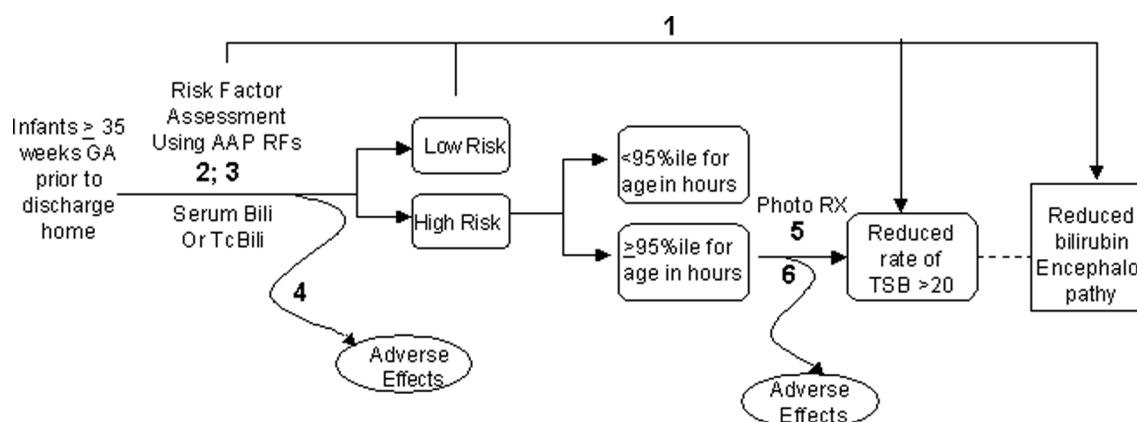
Accepted for publication Mar 11, 2009

Address correspondence to Stanley Ip, MD, Tufts Medical Center, Center for Clinical Evidence Synthesis, Institute for Clinical Research and Health Policy Studies, #63, 800 Washington St, Boston, MA 02111. E-mail: sip@tuftsmedicalcenter.org

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2009 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

**FIGURE 1**

Analytic framework for hyperbilirubinemia screening. AAP indicates American Academy of Pediatrics; RFs, risk factors; Photo RX, phototherapy; %ile, percentile. Key questions: (1) Does screening using risk-factor assessment and/or bilirubin testing reduce the incidence of acute or chronic bilirubin encephalopathy? (2) Does risk-factor assessment accurately identify infants who may benefit from bilirubin testing? (3) Does bilirubin testing accurately identify infants who may benefit from phototherapy? (4) What are the harms of screening? (5) Does treatment reduce the risk of bilirubin encephalopathy in infants identified by screening? (6) What are the harms of treatment with phototherapy? Key questions (5) and (6) are not addressed in this article but are addressed in the Evidence Report.

Some degree of jaundice or hyperbilirubinemia occurs in most newborns. Severe neonatal hyperbilirubinemia is associated with chronic bilirubin encephalopathy or kernicterus, a rare condition characterized by athetoid spasticity, gaze and visual abnormalities, and sensorineural hearing loss in survivors. It may also be associated with mental retardation. In the literature, the term “kernicterus” has been used interchangeably with both the acute and chronic findings of bilirubin encephalopathy. Herein we adopt the suggestions of the American Academy of Pediatrics Subcommittee on Hyperbilirubinemia and reserve the term “kernicterus” for the chronic form of the condition.¹

A 2003 review reported that chronic bilirubin encephalopathy has a mortality rate of at least 10% and morbidity rate of at least 70%.² The true incidence of chronic bilirubin encephalopathy is unknown, because it is not a mandatory reportable disease. A 2001 Joint Commission *Sentinel Event Alert* stated that cases of kernicterus have continued to be reported in recent years.³ There are initiatives to prevent and eliminate this rare disease by in-

stituting widespread screening for hyperbilirubinemia and subsequent timely treatment with phototherapy or exchange transfusion to reduce bilirubin levels.^{1,4,5} Typically, screening pertains to risk stratification on the basis of known risk factors and/or bilirubin-level measurements.

The Tufts Evidence-Based Practice Center (Tufts EPC) completed an evidence report in 2003 examining the effects of bilirubin on neurodevelopmental outcomes in infants of at least 34 weeks' gestation.^{2,6} In 2007, the Center on Primary Care, Prevention and Clinical Partnerships at the Agency for Healthcare Quality and Research (AHRQ), on behalf of the US Preventive Services Task Force (USPSTF), requested that Tufts EPC supplement the 2003 evidence report to address new questions. The findings are used by the USPSTF to make recommendations concerning screening for bilirubin encephalopathy in neonates.

Herein, we summarize key findings from our systematic review that supplements the previous report. We aimed to examine the effects of screening for hyperbilirubinemia on

the incidence of acute and chronic bilirubin encephalopathy by addressing a set of key research questions using evidence-based medicine methodology.

METHODS

Analytic Framework

In a systematic review it is often helpful to develop a schematic (analytic framework) that visually maps the specific linkages that associate the considered populations, interventions, modifying factors, and outcomes.⁷ An analytic framework provides a basis for interpreting and contextualizing relevant studies and establishing which links in the chain of logic have been answered, have inconclusive evidence, or have not yet been addressed. The Tufts EPC, the Center on Primary Care, Prevention and Clinical Partnerships at the AHRQ, and the USPSTF jointly developed an analytic framework and a set of study inclusion/exclusion criteria that are suitable to meet the USPSTF objectives. The analytic framework and the 4 key questions are depicted in Fig 1. This manuscript presents information on the first 4 key questions. Key questions 5

(Does treatment reduce the risk of bilirubin encephalopathy in infants identified by screening?), for which we found no data, and key question 6 (What are the harms of treatment with phototherapy?), for which we found limited data, are presented in the Evidence Report.⁸

Search Strategy and Identification of Relevant Studies

This review answers additional key questions to supplement our previous report. The previous report identified all eligible studies published up to September 2001.^{2,6} We performed additional electronic Medline searches for English-language studies published from September 2001 to August 2007 using Medical Subject Headings (MeSH) terms and key words, such as “jaundice,” “bilirubin,” “hyperbilirubinemia,” and “kernicterus.” We complemented our electronic searches by perusing the reference lists of identified relevant studies and with input from USPSTF lead experts.

Study Eligibility

We included English-language publications on healthy term or preterm infants of at least 35 weeks' gestation. For key questions 1 to 3 (see Fig 1), we considered cohort and (nested) case-control studies on the ability of screening to predict (acute or chronic) bilirubin encephalopathy or clinically significant hyperbilirubinemia. Eligible screening modalities included risk-factor scores, transcutaneous bilirubin (TcB) measurements, early total serum bilirubin (TSB) measurements, or combinations thereof. We also considered reports from US-based cohort studies that described the impact of system-wide implementation of TcB screening and the use of the hour-specific nomogram percentile of jaundice⁴ and (re)admissions for the treatment of hyperbilirubinemia.

We excluded studies on clinical or maternal assessments of jaundice, cord blood bilirubin, end-tidal carbon monoxide, or umbilical cord α -fetoprotein alone. We also rejected studies that compared the agreement between different bilirubin-measurement methods (eg, using correlation analyses or difference versus average analyses [Bland-Altman analyses^{9,10}]) without assessment of diagnostic accuracy.

For key question 4 (harms of screening), we considered all potentially relevant publications including case reports or case series.

Data Extraction

Two experienced reviewers extracted each study in a nonblinded fashion using preconstructed forms. When needed, a third reviewer served as an arbitrator and resolved discrepancies. From each study, we recorded the first author; journal and year of publication; study country and setting; number of infants enrolled and analyzed; study eligibility criteria (including minimum gestational age and birth-weight cutoffs); study design and setting; how study subjects were recruited (consecutively, by random sampling, or otherwise); and information on the characteristics of the screening test and the reference standard for hyperbilirubinemia. For TcB measurements, we recorded the device used and timing of measurements. For risk instruments, we recorded the components in the composite score. We also extracted sensitivities and specificities for various reported cutoffs of TcB or early TSB measurements (as applicable) and noted area-under-the-curve (AUC) values (see below).

Assessment of Methodologic Quality

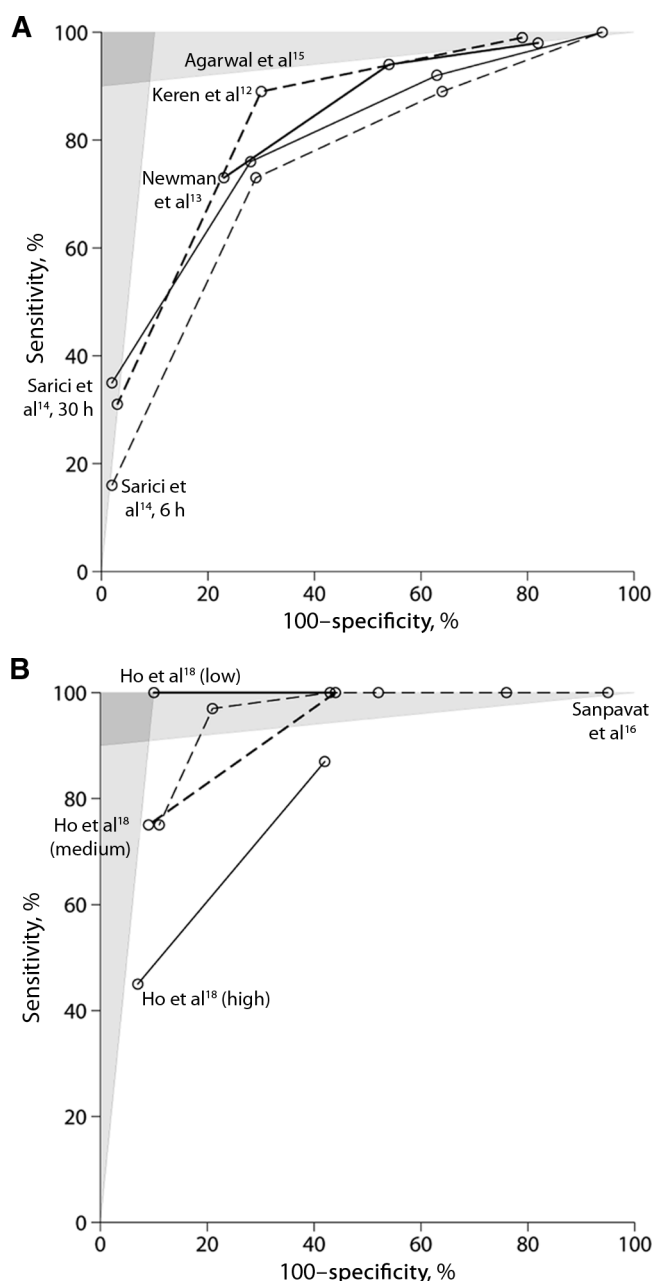
Each study was assigned a quality rating of “good,” “fair,” or “poor” by 2 reviewers, per USPSTF criteria,⁷ and the presence or absence of overt method-

ologic errors.¹¹ When needed, a third reviewer acted as an arbitrator. Briefly, we assessed study characteristics commonly associated with less susceptibility to biases and systematic errors, such as clarity of outcome definitions, suitability of statistical methods, and proper accounting for confounders.⁷ We considered the presence of overt verification bias as an important methodologic error. Verification bias operates when only infants with positive screening test results (eg, high scores in the risk instruments or high TcB measurements) were verified with the reference standard (late serum TSB measurements), whereas infants with negative screening results (eg, low scores or TcB measurements) were not. This can result in upward-biased sensitivity and downward-biased specificity estimates. We based our qualitative conclusions on studies of good or fair quality.

Data Analysis and Presentation

Sensitivity, Specificity, and AUC

Because of dissimilarities in the identified studies, no quantitative synthesis (meta-analysis) was performed. However, we calculated from each study the corresponding sensitivity and specificity pairs and depict them in sensitivity/100%-specificity plots. The sensitivity of a screening test is its ability to maximize true-positive diagnoses, and specificity is its ability to minimize false-positive diagnoses. A perfect screening test has both sensitivity and specificity equal to 100%. There is a trade-off between the sensitivity and specificity of a test. As the cutoff for a positive screening test decreases (eg, the value of TcB above which measurements are considered suggestive of a high bilirubin level), the corresponding sensitivity increases and the corresponding specificity decreases. One can capture this trade-off by recording multiple sensitivity/

**FIGURE 2**

Diagnostic ability of early TSB and TcB measurements. A, Ability of early TSB measurements to identify postdischarge TSB above the 95th hour-specific percentile. B, Ability of TcB measurements to identify high TSB (as defined in Table 3). Sensitivity/specificity pairs from the same study (obtained with different cutoffs for the early TSB measurement) are connected with lines. These lines may not be representative of the ROC curve of the pertinent studies. Studies listed on the left shaded area have an LR^+ of ≥ 10 ; studies listed on the top shaded area have an LR^- of ≤ 0.1 ; and studies listed at the intersection of the gray areas (darker gray polygon) have both an LR^+ of >10 and an LR^- of <0.1 .

specificity pairs corresponding to different cutoffs. These can be plotted in a square sensitivity versus 100%-specificity plot (ie, constructing a receiver operating characteristic curve [see Fig 2 for examples]). The area under

the (receiver operating characteristic) curve can summarize diagnostic ability across all positivity cutoffs. AUC values of 0.5 imply lack of any diagnostic ability, whereas AUC values of 1.0 correspond to a perfect screening test.

Positive and Negative Likelihood Ratios

We also characterized diagnostic abilities by using positive and negative likelihood ratios (LR^+ and LR^- , respectively). These quantities express the strength of the diagnostic or predictive information conveyed by the screening test results. Briefly, LR^+ expresses how much a positive screening test reinforces our pre-screening belief that an infant will indeed develop a high bilirubin level. Conversely, LR^- quantifies the extent to which a negative screening result reinforces our prescreening belief that an infant will indeed not develop a high bilirubin level. LR^+ and LR^- values of 1 imply no diagnostic ability (ie, our pre-screening belief is multiplied by a factor of 1 and, therefore, remains unaltered). As suggested in the literature, we consider tests with either an LR^+ of >10 or an LR^- of <0.1 informative and clinically useful.¹² Instead of providing tables of LRs per study and cutoff used, we incorporated this information in the graphical analyses.

Software

Intercooled Stata 8.2 (Stata Corp, College Station, TX) was used for calculations and graphics.

RESULTS

We screened 742 abstracts and 96 articles qualified for full-text examination, 9 of which were deemed eligible for the current systematic review.^{4,13–20} We show a schematic of the screening process in Appendix 1. In addition, we included a single study that was identified in our previous evidence report.²¹

Key Question 1: Does Screening Using Risk-Factor Assessment and/or Bilirubin Testing Reduce the Incidence of Acute or Chronic Bilirubin Encephalopathy?

No study directly addressed this question. Eligible studies evaluated only

TABLE 1 Summary of Studies That Used Risk Predictive Instruments to Predict High TSB Measurements

Author (Year), Country	Design	N (Cases)	BW, g	GA, wk	High TSB Levels	Verification Bias	Validation	AUC, 95% CI	Cutoff ^a
Keren et al ¹³ (2005), United States	Retrospective cohort (SC)	993 (NA)	≥2000 ^b	≥36	>95th percentile (postdischarge)	No	No	0.71 (0.66–0.76)	≥12
Newman et al ¹⁴ (2005) (study 2), United States	Retrospective cohort (MC)	5706 (NA)	>2000	>36	>20 (by 48 h)	No	No	0.69 (—)	≥10
Newman et al ¹⁴ (2005) (study 1), United States	Nested case-control (MC)	275 (67)	2000	>36	>25 (by 30 d)	No	Yes ^c	0.83 (0.77–0.89)	—
Newman et al ²¹ (2000), United States	Nested case-control (MC)	496 (73) ^d	>2000	>36	>25 (by 48 h)	No	No	0.84 (0.79–0.89)	≥10

All studies received “fair” grading for their methodologic quality. BW indicates birth weight; CI, confidence interval; GA, gestational age; MC, multicenter; NA, not applicable; SC, single center.

^a Cutoff for the risk instrument (see Appendix for details) suggested by the study as the 1 with the best predictive ability.

^b At or higher than 2500 g if ≥35 weeks’ gestation.

^c Independent validation of the score developed by Newman et al.¹³ However, the score was “modified,” abolishing 1 of the most influential factors (family history of jaundice in newborn) because it was available for only 10% of the infants.

^d All cases of significant hyperbilirubinemia in the second 2005 Newman et al¹³ study are probably included in their first 2005 study¹³ and in their 2000 study.²⁰

surrogate outcomes, namely the incidence of hyperbilirubinemia.

Key Question 2: Does Risk-Factor Assessment Accurately Identify Infants Who May Benefit From Bilirubin Testing?

We presumed that infants who would benefit from phototherapy are those with high bilirubin levels. Four studies described in 3 publications examined the effectiveness of 2 risk instruments assessing infant and family history in screening for the development of significant hyperbilirubinemia (Table 1), all of which were conducted in the United States. Two were retrospective cohorts,^{13,14} and 2 were nested case-control studies.^{14,21} All studies enrolled

infants with available information on factors included in the risk instruments and also TSB measurements before and after discharge,¹³ anytime within 30 days from birth (nested case-control study¹⁴) or after the first 48 hours of life (retrospective cohort¹⁴). The definition of clinically significant hyperbilirubinemia varied across studies (Table 1). All studies were graded as being of fair quality.

The 2 studied risk instruments had only 2 risk factors in common (of 6¹³ and 9^{14,21}), namely exclusive breastfeeding and gestational age (Appendix 2). However, even these common factors contributed differently to the total risk score/index. Therefore, the same

infant can receive different risk scores with different instruments.

Overall, evidence suggests comparable ability of the 2 risk instruments in predicting later significant hyperbilirubinemia (Table 1). Studies reported AUC values ranging from 0.71 to 0.84, with nested case-control studies showing slightly higher AUC values when compared with retrospective cohorts.

Key Question 3: Does Bilirubin Testing Accurately Identify Infants Who May Benefit From Phototherapy?

We discuss separately early measurements of serum bilirubin (early TSB)

TABLE 2 Summary of Studies That Used Early TSB Measurements to Predict High Late TSB Measurements

Author (Year), Country	Design	N _A (N _E)	BW, g	GA, wk	Early TSB		Late TSB		Verification Bias	AUC (95% CI)
					Timing	Cutoff, Percentile ^a	Timing, h	Cutoff, Percentile ^a		
Keren et al ¹³ (2005), United States (SC)	RC	996 (899) ^b	≥2000 ^c	≥36	Pre-discharge	40, 75, 95	Postdischarge	95	No	0.83 (0.80–0.86)
Newman et al ¹⁴ (2005), United States (MC)	RC	5711 (5706)	>2000	>36	<48 h	40, 75, 95	≥48 h	>95 (>20 mg/dL)	No	0.83 (0.80–0.85)
Sarici et al ¹⁵ (2004), Turkey (SC)	PC	366 (146) ^d	—	35–42	6 and 30 h	5, 30, 60, 95	4 times until 150 h	95	No	—
Agarwal et al ¹⁶ (2002), India (SC)	PC	220 (213)	—	≥35	24 h	≈75 (6 mg/dL)	72 h	95	Yes ^e	—

BW indicates birth weight; CI, confidence interval; GA, gestational age; MC, multicenter; N_A (N_E), number analyzed (number enrolled); PC, prospective cohort; RC, retrospective cohort; SC, single center.

^a Percentile in the Bhutani et al⁴ hour-specific nomogram.

^b These infants were included among those analyzed in development of the Bhutani et al⁴ nomogram for hour-specific TSB values.

^c At or higher than 2500 g if ≥35 weeks’ gestation.

^d Only 146 near-term infants were analyzed (not stated why).

^e Late TSB was measured only among those with >10 mg/dL (unclear how this was deduced); 69% of the infants were verified.

TABLE 3 Summary of Studies That Used TcB to Identify Infants With TSB Suggesting Phototherapy

Author (Year), Country	Design	N_E (N_A)	BW, g	GA, wk	Timing	TcB				Definition of High TSB Level	Verification Bias
						Device	Site	Cutoff, Percentile ^a	Δt From TSB		
Ho et al ¹⁹ (2006), Hong Kong, China (SC)	RC	1621 (997) ^b	—	>35	First 3 d	JM-103	Midsternum	75, 95	30 min	3 different cutoffs: high, medium, and low risk ^c	No
Sanpavat et al ¹⁷ (2005), Thailand (SC)	Unclear	392 (392)	≥2000	≥37	—	BiliCheck	Forehead	10, 25, 50, 75, 85, 90, 95	—	25–48 h: ≥10 mg/dL; 49–72 h: ≥13 mg/dL; >72 h: ≥15 mg/dL	Yes ^d

Δt from TSB indicates time interval between the 2 measurements; N_A (N_E), number analyzed (number enrolled); SC, single center; RC, retrospective cohort. Devices: JM-103 (Minolta Air-Shields, Chuo-Ku, Osaka, Japan) and BiliCheck (SpectRx, Inc, Norcross, GA).

^a Percentile in the Bhutani et al⁴ hour-specific nomogram.

^b A total of 997 had TcB measurements above the 40th hour-specific percentile and available paired TSB measurements.

^c Infants at low risk: ≥ 38 weeks' gestation and well; infants at medium risk: ≥38 weeks' gestation with risk factors or 35 to 37 weeks' gestation and well; infants at high risk: 35 to 37 weeks' gestation with risk factors.

^d Only 39% ($n = 154$) of the infants received TSB measurements (unclear how these were selected ["jaundice"]).

and TcB measurements. Again, we presumed that infants who would benefit from phototherapy are those with high bilirubin levels.

Early TSB Measurements

Four studies were eligible (Table 2).^{13–16} One prospective study from India¹⁶ included all 220 healthy infants with a gestational age of ≥35 weeks born within a 5-month period. This study verified late TSB values only for infants with “at least 10 mg/dL” (ie, had overt verification bias). The other prospective study from Turkey¹⁵ included 366 consecutive infants with a gestational age between 35 and 42 weeks but reported data on the diagnostic ability of early TSB measurements only among 146 infants with a gestational age between 35 and 37 weeks. The 2 retrospective studies from the United States^{13,14} were described in key question 2. In all studies, the reference standard was a postdischarge measurement above the hour-specific 95th percentile. Three studies received a fair grade^{13–15} and 1 received a poor grade¹⁶ for their methodologic quality. Fig 2A illustrates the diagnostic ability of early measurements to identify postdischarge TSB levels above the 95th hour-specific percentile. All 4 studies reported comparable diagnostic abilities (Table 2 and Fig 2A). Only

the study with overt verification bias¹⁶ had an LR^- of <0.1 for an early TSB cutoff of 6 mg/dL during the first 24 hours (which is near the 75th hour-specific percentile in the Bhutani et al nomogram⁴).

Combination of Risk Instruments and Early TSB Levels

In a retrospective cohort,¹⁴ the effectiveness of a combination of a risk instrument (see Appendix 2 for details) with early TSB measurements in predicting a TSB level of ≥20 mg/dL at ≥48 hours after birth (ie, levels above the 95th hour-specific percentile) was evaluated. The AUC value improved from 0.69 to 0.86 ($P < .05$) after incorporating the z scores of TSB measurements in the predictive model (the z scores express how extreme a TSB measurement is, in SD units).

TcB Measurements

One study from Thailand¹⁷ and 1 study from China (Hong Kong)¹⁹ were eligible; both studies included almost exclusively infants of Asian descent. Both studies analyzed selected infant subgroups on the basis of availability of measurements, and 1 of them had overt verification bias. The studies defined the presence of high bilirubin levels in different ways (Table 3). Both were graded as poor for their method-

ologic quality. Fig 2B summarizes the diagnostic ability of TcB measurements to identify a TSB level that was indicative of phototherapy in these 2 studies.

Outcomes After the Implementation of Screening Strategies

Three retrospective studies compared rates of phototherapy or readmission for hyperbilirubinemia before and after the implementation of screening (Table 4).^{4,18,20} All were descriptive studies without concurrent controls and without adjustments for confounders. All received a poor grade for their methodologic quality.

The first study⁴ described the implementation of a “systems approach” in 3 sequential incremental steps (Table 4). In the first step, nurses were authorized to obtain TSB or TcB measurements for clinical jaundice. In the second step, all infants received pre-discharge TSB measurements (at routine metabolic screening). During this phase, the Bhutani et al hour-specific bilirubin nomogram was developed.⁴ Finally, in the third step, the hour-specific nomogram was used to interpret the universal TSB measurements at discharge. Intensive phototherapy rates (in nursery and after readmissions) increased from 4.5% (first step) to 5.4% (second step) and

TABLE 4 Outcomes Associated With the Implementation of Bilirubin Screening in Term and Near-Term Infants

Author (Year), Country	Period	Screening Strategy		Enrolled, <i>n</i>	Phototherapy, %	Exchange Transfusion, <i>n</i>	Mean LOS, d	Readmission Rate ^a
		Description	Interpretation					
Bhutani et al ⁴ (2006), United States	1993–1995	No screening ^b	—	7929	4.5	6	—	14
	1996–1998	TSB at discharge (all)	Emp (?)	8186	5.4	5	—	11
	1999–2000	TSB at discharge (all)	BN	6395	2.5	2	—	5.5
Eggert et al ²⁰ (2006), United States	2001–2002	No screening	—	48 798	—	0	—	5.5
	2003–2004	TSB or TcB at discharge (all) ^c	modBN ^c	52 483	—	0	—	4.3
Petersen et al ¹⁸ (2005), United States	2002–2003	No screening	—	[405] ^d	[5.9] ^d	—	2.2	4.5
	2003	TcB at discharge (all)	BN	[421] ^d	[7.7] ^d	—	2.1	1.8

(mod)BN indicates (modified) Bhutani et al⁴ nomogram to interpret bilirubin measurements; Emp (?), empirical interpretation of bilirubin measurements (unclear); LOS, length of stay.

^a For hyperbilirubinemia or phototherapy, per 1000 discharges.

^b Nurses were authorized to perform TSB/TcB measurements for clinical jaundice.

^c The care provider of infants with TSB (or TcB) levels above the 40th percentile on the modified nomogram was notified and decided on further management.

^d These numbers express average rates per month during the duration of the study, rather than the total number of enrolled infants or total percentage receiving phototherapy, as applicable.

then decreased to 2.5% during the third step. Readmission rates gradually decreased throughout the study.

The report of another study described the implementation of universal pre-discharge TSB or TcB screening in 18 hospitals.¹⁹ After the implementation of screening, the proportion of term and near-term infants with “severe hyperbilirubinemia” increased significantly compared with the baseline. However, the proportion of readmissions decreased significantly ($P < .005$). Authors of the third study¹⁸ reported that the proportion of newborns treated with phototherapy while in the nursery increased significantly after initiation of TcB screening compared with baseline rates. The mean rate of readmission for hyperbilirubinemia decreased significantly over time (Table 4; $P = .044$).

Key Question 4: What Are the Harms of Screening?

None of the reviewed studies assessed the harms of screening.

DISCUSSION

We supplemented our previous evidence report² by summarizing the evidence on hyperbilirubinemia screening. The evidence presented in this review will be used by the USPSTF to make recommendation statements on screening for neonatal hyperbiliru-

binemia. No study evaluated the effects of screening on the rates of acute or chronic bilirubin encephalopathy. Instead, studies evaluated the surrogate outcome of bilirubin encephalopathy or hyperbilirubinemia (adopting different definitions). Overall, screening by use of risk factors, TcB or early TSB measurements, or combinations thereof is effective in predicting a high bilirubin level. Descriptive uncontrolled evidence suggests that screening is associated with increased diagnoses of hyperbilirubinemia and fewer readmissions for hyperbilirubinemia. All aforementioned findings are based on studies with methodologic problems and pragmatic limitations.

Study of the analytic framework in Fig 1 can help contextualize findings. We find that screening can predict high bilirubin levels. However, to connect this finding to clinical outcomes, we have to make the assumption that a high bilirubin level is a valid surrogate for bilirubin encephalopathy.²² This means that, foremost, a high bilirubin level must be in the causal pathway between screening (intervention) and bilirubin encephalopathy (clinical outcome). This is generally supported by the literature. Second, we make the assumption that if screening does affect the rates of bilirubin encephalopathy, it would do so by treatments implemented as a result of the screening.

This implies that any intervention that lowers high bilirubin levels will also prevent bilirubin encephalopathy. However, we have no explicit data to show that treatments such as phototherapy or exchange transfusion actually decrease the risk of bilirubin encephalopathy. Despite the lack of data, it is generally postulated that hyperbilirubinemia is a valid surrogate outcome for acute and chronic bilirubin encephalopathy. A final caveat is that, even accepting hyperbilirubinemia as a valid surrogate end point, the diagnostic accuracy of the studied screening modalities is based on relatively few studies, some of which have overt methodologic problems.

We also found tangential evidence from cohort studies that screening is associated with increased detection of hyperbilirubinemia and fewer readmissions for hyperbilirubinemia. Notwithstanding reservations on the validity of these intermediate clinical outcomes, we caution that they were observed in studies without concurrent controls and without analyses that can account for confounding and secular trends. Therefore, one cannot be confident that the reported favorable changes in the intermediate clinical outcomes are attributable solely to the initiation of system-wide screening.

It is generally difficult to appreciate the trade-off in the effectiveness, harms, and costs of screening without performing a formal decision analysis.²³ This is especially true for rare conditions, such as chronic bilirubin encephalopathy, and whenever key parameters are unknown (such as the effectiveness of screening strategies to prevent the disease). Allowing for these caveats, a recent cost-effectiveness analysis concluded that widespread screening for bilirubin encephalopathy is probably going to increase health costs substantially, with uncertain benefits.²⁴

There are several limitations to this work. First, this is a supplement to our previous evidence report; we used a different literature search strategy and narrower study eligibility criteria. Therefore, one may argue that because we relied on the previous evidence report we may have missed eli-

gible articles that were published before 2001. However, all eligible articles for the current review would also be eligible for the broader key questions of the original evidence report. Second, the number of eligible studies that addressed the key questions was limited, and none were graded as having good methodologic quality. Finally, because we had few studies and they were clinically and methodologically heterogeneous, we did not perform quantitative analyses.

CONCLUSIONS

In principle, a definitive answer to the question of whether screening for hyperbilirubinemia can reduce the rate of acute and chronic encephalopathy would be conveyed by an adequately powered, pragmatic (cluster) randomized trial. Because kernicterus is rare (0.9 cases per 100 000 live births in a prospective study in the United Kingdom and Ireland²⁵), such a trial

would be challenging to perform. For practical consideration, studies on the effectiveness of different strategies to reduce the incidence of bilirubin encephalopathy could only rely on a surrogate outcome such as hyperbilirubinemia. Because severe hyperbilirubinemia is a rare event²⁵ (and because of the clustered design), tens of thousands of infants per arm would be needed to attain statistical power, posing questions on the feasibility of such a study. Future studies on preventive and screening strategies should also actively monitor the potential harms from implementing such strategies in both infants and their family members.

ACKNOWLEDGMENTS

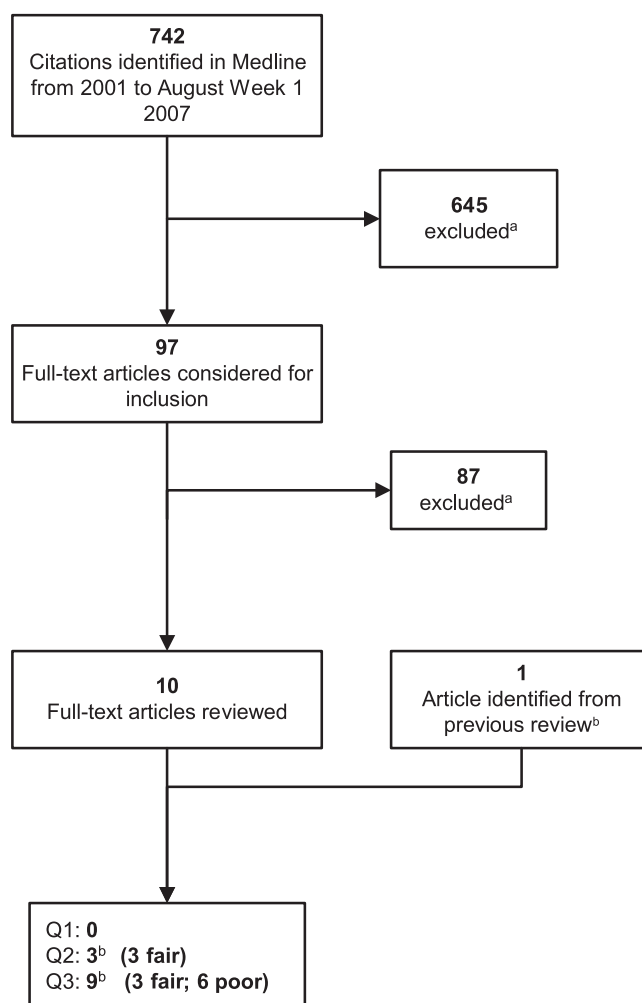
This article is based on a report conducted by the Tufts EPC under contract to the AHRQ (contract 290-020022).

We thank Ms Audrey Mahoney for administrative support.

REFERENCES

1. American Academy of Pediatrics, Subcommittee on Hyperbilirubinemia. Management of hyperbilirubinemia in the newborn infant 35 or more weeks of gestation [published correction appears in *Pediatrics*. 2004;114(4):1138]. *Pediatrics*. 2004;114(1):297–316
2. Ip S, Glick S, Kulig J, O'Brien R, Sege R. Management of neonatal hyperbilirubinemia. *Evid Rep Technol Assess (Summ)*. 2003;(65):1–5
3. Kernicterus threatens healthy newborns. *Sentinel Event Alert*. 2001;(18):1–4
4. Bhutani VK, Johnson LH, Schwoebel A, Gennaro S. A systems approach for neonatal hyperbilirubinemia in term and near-term newborns. *J Obstet Gynecol Neonatal Nurs*. 2006;35(4):444–455
5. Chou SC, Palmer RH, Ezhuthachan S, et al. Management of hyperbilirubinemia in newborns: measuring performance by using a benchmarking model. *Pediatrics*. 2003;112(6 pt 1):1264–1273
6. Ip S, Chung M, Kulig J, et al; American Academy of Pediatrics, Subcommittee on Hyperbilirubinemia. An evidence-based review of important issues concerning neonatal hyperbilirubinemia. *Pediatrics*. 2004;114(1). Available at: www.pediatrics.org/cgi/content/full/114/1/e130
7. Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med*. 2001;20(3 suppl):21–35
8. Ip S, Chung M, Trikalinos T, DeVine D, Lau J. Screening for Bilirubin Encephalopathy. Evidence Synthesis No. 72. AHRQ Publication No. 09-05136-EF-1. Rockville, Maryland: Agency for Healthcare Research and Quality, October 2009
9. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310
10. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8(2):135–160
11. Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuyt PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess*. 2004;8(25):iii, 1–234
12. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA*. 1994;271(9):703–707

13. Keren R, Bhutani VK, Luan X, Nihtianova S, Cnaan A, Schwartz JS. Identifying newborns at risk of significant hyperbilirubinaemia: a comparison of two recommended approaches. *Arch Dis Child*. 2005;90(4):415–421
14. Newman TB, Liljestrand P, Escobar GJ. Combining clinical risk factors with serum bilirubin levels to predict hyperbilirubinemia in newborns. *Arch Pediatr Adolesc Med*. 2005;159(2):113–119
15. Sarici SU, Serdar MA, Korkmaz A, et al. Incidence, course, and prediction of hyperbilirubinemia in near-term and term newborns. *Pediatrics*. 2004;113(4):775–780
16. Agarwal R, Kaushal M, Aggarwal R, Paul VK, Deorari AK. Early neonatal hyperbilirubinemia using first day serum bilirubin level. *Indian Pediatr*. 2002;39(8):724–730
17. Sanpavat S, Nuchprayoon I, Smathakane C, Hansuebsai R. Nomogram for prediction of the risk of neonatal hyperbilirubinemia, using transcutaneous bilirubin. *J Med Assoc Thai*. 2005;88(9):1187–1193
18. Petersen JR, Okorodudu AO, Mohammad AA, Fernando A, Shattuck KE. Association of transcutaneous bilirubin testing in hospital with decreased readmission rate for hyperbilirubinemia. *Clin Chem*. 2005;51(3):540–544
19. Ho HT, Ng TK, Tsui KC, Lo YC. Evaluation of a new transcutaneous bilirubinometer in Chinese newborns. *Arch Dis Child Fetal Neonatal Ed*. 2006;91(6):F434–F438
20. Eggert LD, Wiedmeier SE, Wilson J, Christensen RD. The effect of instituting a prehospital-discharge newborn bilirubin screening program in an 18-hospital health system. *Pediatrics*. 2006;117(5). Available at: www.pediatrics.org/cgi/content/full/117/5/e855
21. Newman TB, Xiong B, Gonzales VM, Escobar GJ. Prediction and prevention of extreme neonatal hyperbilirubinemia in a mature health maintenance organization. *Arch Pediatr Adolesc Med*. 2000;154(11):1140–1147
22. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. *JAMA*. 1999;282(8):771–778
23. Pauker SG. Deciding about screening. *Ann Intern Med*. 1993;118(11):901–902
24. Suresh GK, Clark RE. Cost-effectiveness of strategies that are intended to prevent kernicterus in newborn infants. *Pediatrics*. 2004;114(4):917–924
25. Manning D, Todd P, Maxwell M, Jane PM. Prospective surveillance study of severe hyperbilirubinaemia in the newborn in the UK and Ireland. *Arch Dis Child Fetal Neonatal Ed*. 2007;92(5):F342–F346

**APPENDIX 1**

Study eligibility flowchart. ^a Management of neonatal hyperbilirubinemia²; ^b 2 articles provided data for both questions 2 and 3. *Qn* indicates question number.

APPENDIX 2 Comparisons of the Risk Scores/Indexes in the 3 Studies That Were Eligible for Key Question 2

Risk Factors	Weighting for Risk Score/Index		
	Keren et al ¹² (2005)	Newman et al ¹³ (2005)	Newman et al ²⁰ (2000)
Exclusive breastfeeding	5	6	6
Breastfeeding and bottle feeding	4	—	—
Family history of jaundiced newborns	—	—	6
Bruising	—	4	4
Asian race	—	4	4
Cephalohematoma	—	3	3
Mother age ≥ 25 y	—	3	3
Male gender	—	1	1
Black race	—	−2	−2
Gestational age			
40 wk	—	2	2
<38 wk	5	—	—
3 points per 500 g above 2000 g in birth weight ^a	3	—	—
Oxytocin used	4	—	—
Vacuum delivery	4	—	—

^a Calculation: (birth weight [g] − 2001)/500, rounding down to the nearest integer.