



Management of Neonatal Hyperbilirubinemia

Summary

Overview

The report from which this summary was developed presents a comprehensive literature review of the effect of bilirubin on neurodevelopmental outcomes. It also examines the role of various effect modifiers such as sepsis and hemolysis on neurodevelopment, the efficacy of phototherapy, the accuracy of transcutaneous measurement of bilirubin, and the various strategies in predicting hyperbilirubinemia.

As background, in 1994, the American Academy of Pediatrics (AAP) published guidelines on the management of neonatal hyperbilirubinemia developed by the Provisional Committee for Quality Improvement and Subcommittee on Hyperbilirubinemia. Although there was no definitive evidence showing a specific level of bilirubin and subsequent serious adverse neurodevelopmental outcome, the task force, relying on retrospective epidemiologic data primarily derived from North American and European research, offered recommendations for the management of neonatal hyperbilirubinemia. These were based on evidence when appropriate data existed, and based on expert consensus when data were lacking.

These recommendations were specifically directed at evaluation and treatment of hyperbilirubinemia in healthy term newborns; i.e., only infants without signs of illness or apparent hemolytic disease. Highlights in the recommendations included visual inspection of the skin to determine jaundice, use of total serum bilirubin (TSB) level as the relevant variable in determining treatment, and recommendation for exchange transfusion only if intensive phototherapy fails to lower the TSB to less than

20 mg/dl. Criticisms regarding these recommendations included the fact that:

- Visual inspection of jaundice may not be reliable.
- Criteria based on age measured by number of days rather than hours are too coarse for precise interpretation of bilirubin level.
- Evidence about whether there is any adverse effect of phototherapy treatment on healthy term newborns is absent.

Furthermore, some term infants without evidence of hemolysis may develop hyperbilirubinemia and kernicterus, and there have not been any randomized controlled trials to assess the relation between bilirubin levels and adverse neurodevelopmental effects. Through review of evidence for five key questions, the report aims to supply data for an update of these recommendations.

Reporting the Evidence

The Evidence-based Practice Center (EPC) formed an evidence review team consisting of pediatricians and EPC methodological staff to review the literature and perform data abstraction and analysis. The evidence review team held meetings and teleconferences with external technical experts representing the AAP, the American Academy of Family Physicians, the National Association of Pediatric Nurse Practitioners, the Center for Quality of Care Research and Education, the Harvard School of Public Health, and the Parents of Infants and Children with Kernicterus organization. The EPC and its panel of external technical experts refined key questions proposed by the AAP and identified issues central to this report. A comprehensive search of the medical literature was conducted to



identify the evidence available to address the following questions:

Association of Neonatal Hyperbilirubinemia With Neurodevelopmental Outcomes

1. *What is the relationship between peak bilirubin levels and/or duration of hyperbilirubinemia and neurodevelopmental outcome?*
2. *What is the evidence for effect modification of the results in Question 1, by gestational age, hemolysis, serum albumin, and other factors?*

Treatments for Neonatal Hyperbilirubinemia

3. *What are the quantitative estimates of efficacy of treatment at (1) reducing peak bilirubin levels (e.g., number needed to treat at 20 mg/dl to keep TSB from rising); (2) reducing the duration of hyperbilirubinemia (e.g., average number of hours by which time TSB greater than 20 mg/dl may be shortened by treatment); and (3) improving neurodevelopmental outcomes?*

Diagnosis of Neonatal Hyperbilirubinemia

4. *What is the efficacy of various strategies for predicting hyperbilirubinemia, including hour-specific bilirubin percentiles?*
5. *What is the accuracy of transcutaneous bilirubin (TcB) measurements?*

Methodology

Patient Population and Settings

The target population included infants of at least 34 weeks gestational age. Based on the findings of an earlier National Institute of Child Health and Human Development (NICHD) study, in which none of the 1,339 infants greater than or equal to 2,500 grams was less than 34 weeks, the EPC grouped infants weighing greater than or equal to 2,500 grams with those greater than or equal to 34 weeks gestation.

Literature Search and Review Parameters

Search Strategies. MEDLINE® and PreMEDLINE® databases were searched for the evidence report. In September 2001, the MEDLINE® database was searched for publications from 1966 to the present using relevant MeSH terms (“hyperbilirubinemia,” “hyperbilirubinemia, hereditary,” “bilirubin,” “jaundice, neonatal,” “kernicterus”) and text words (“bilirubin,” “hyperbilirubin\$,” “jaundice,” “kernicterus,” “neonat\$”). The abstracts were limited to human and English studies focusing on newborns between birth and 1 month of age. In addition, the same text words used for the MEDLINE® search were used to search the PreMEDLINE® database. The strategy yielded 4,280 MEDLINE® and 45

PreMEDLINE® abstracts. The EPC consulted domain experts and examined relevant review articles for additional studies.

Study Selection. Preliminary screening of abstracts for each question identified over 600 potentially relevant articles for Questions 1, 2, and 3. For Questions 1 and 2, only studies that reported neurodevelopmental outcomes were included. Except for part of Question 3, studies concerning effects of different variables on bilirubin without neurodevelopmental outcome were not included in this review. For the specific question of quantitative estimates of treatment efficacy, all studies concerning therapies for bilirubin ≥ 20 mg/dl were included in the review. The inclusion and exclusion criteria for the systematic review were discussed in several teleconferences of the EPC evidence review team and technical experts. The criteria underwent several revisions before final acceptance by the panel members. The final screening criteria for inclusion and exclusion of articles are described below.

Association of Neonatal Hyperbilirubinemia With Neurodevelopmental Outcomes

For the two questions on the association of neonatal hyperbilirubinemia with neurodevelopment outcomes, the inclusion criteria were: Infants ≥ 34 weeks of gestation or $\geq 2,500$ grams and a sample size of more than five subjects per arm (except for case reports of kernicterus). The predictors were jaundice or hyperbilirubinemia and at least one of the neurodevelopmental outcomes was reported in the article. The study designs included prospective cohorts (more than two arms), prospective cross-sectional study, prospective longitudinal study, prospective single-arm study, or retrospective cohorts (more than two arms).

Treatments for Neonatal Hyperbilirubinemia

For Question 3 on the treatments for neonatal hyperbilirubinemia, studies which focused on the number needed to treat had the following inclusion criteria: Infants ≥ 34 weeks of gestation or $\geq 2,500$ grams and a study size of more than 10 subjects per arm. Treatments included any treatment for neonatal hyperbilirubinemia. Outcomes included serum bilirubin level ≥ 20 mg/dl or frequency of exchange transfusion specifically for bilirubin level ≥ 20 mg/dl. The study design was randomized or non-randomized controlled trials.

For all other studies reviewed for Question 3, the selection criteria were: Infants ≥ 34 weeks of gestation or $\geq 2,500$ grams and sample size of more than 10 subjects per arm for phototherapy; any sample size for other treatments. Any treatment for neonatal hyperbilirubinemia was included and at least one neurodevelopmental outcome was reported in the article.

Diagnosis of Neonatal Hyperbilirubinemia

For Questions 4 or 5 on the diagnosis of hyperbilirubinemia, the inclusion criteria were: Infants ≥ 34 weeks of gestation or birthweight $\geq 2,500$ grams and a sample size of more than 10

subjects. The reference standard was laboratory-based serum bilirubin.

Results of Abstract and Article Screening

Six hundred and sixty-three of a total 4,560 abstracts were identified as potentially relevant articles after preliminary screening. There were 158, 174, 99, 153, and 79 abstracts for Questions 1, 2, 3, 4, and 5, respectively.

After full-text screening (according to the inclusion and exclusion criteria described above), 138 of 253 retrieved articles were included in this report. Twenty-eight articles reported on cases of kernicterus, 35 articles reported correlations, 21 articles reported on treatments, and 54 articles were included in the review of diagnosis. There was some inevitable overlap because treatment effects and neurodevelopmental outcomes were inherent in the study designs.

Methodological Quality

Methodological quality or internal validity addresses the design, conduct, and reporting of the study. Some of the items belonging to this domain are widely used in various “quality” scales and for randomized controlled trials, and usually include items such as concealment of random allocation, treatment blinding, and handling of dropouts. Because different types of study designs are used to address different questions and for consistency in the interpretation across different designs, a three-category scale was defined to report the methodological quality of the studies in the evidence report: A (least bias), B (susceptible to some bias), or C (likely to have large bias).

Studies of Association (Questions 1 and 2). The criteria for evaluating methodological quality of studies that assess association are:

- A—Prospective. Complete methods and results (including inclusion/exclusion criteria). Proper control/comparison group, correct analyses performed.
- B—Prospective or retrospective. Not all criteria of A. Some deficiencies; however, unlikely to cause major bias.
- C—Prospective or retrospective. Significant design or reporting errors, large amount of missing information or bias.

Studies of Treatments (Question 3). The criteria for evaluating methodological quality of studies that assess effects of treatments are:

- A—Randomized controlled trial. Complete methods and results (including inclusion/exclusion criteria) described. Proper randomization and/or blinding, and correct analyses performed.
- B—Non-randomized controlled trial or other prospective design (prospective cohort or case-control study). Proper

selection of control group. Not all criteria of A. Some deficiencies; however, unlikely to cause major bias.

- C—Retrospective or no control group. Significant design or reporting errors, large amount of missing information or bias.

Studies of Diagnosis (Questions 4 and 5). The criteria for evaluating methodological quality of studies that assess diagnostic test performance are:

- A—Prospective. Complete methods and results (including inclusion/exclusion criteria) described. Proper reference standard used and correct analyses performed.
- B—Prospective or retrospective. Not all criteria of A. Some deficiencies; however, unlikely to cause major bias.
- C—Prospective or retrospective. Significant design or reporting errors, large amount of missing information or bias.

Statistical Analysis

The number needed to treat (NNT), expressing the benefit of an active treatment over a control, was calculated to quantify the efficacy of treatment for neonatal hyperbilirubinemia. For Question 3 in this report, NNT can be interpreted as the number of newborns needed to be treated at 20 mg/dl to keep the TSB in one newborn from rising.

A meta-analysis of Question 4 was conducted using the summary receiver operating characteristic (ROC) method to combine studies which evaluated diagnostic test performance. A meta-analysis of correlation coefficients was conducted to correlate performance of transcutaneous bilirubin measurements with serum bilirubin.

Results

Question 1

What is the relationship between peak bilirubin levels and/or duration of hyperbilirubinemia and neurodevelopmental outcome?

- A summary of 28 reports that spanned over 30 years of 123 cases of kernicterus in term/near-term infants affirms the role of elevated bilirubin level in kernicterus. The disease, although infrequent, has significant mortality (at least 10 percent) and long-term morbidity (at least 70 percent). It is important to note that a significant amount of demographic information was missing in these case reports. Thirty-five term/near-term (≥ 34 weeks gestation) infants with idiopathic hyperbilirubinemia developed kernicterus with a TSB level ranging from 22.5 mg/dl to 54 mg/dl. Eighty-eight infants with hyperbilirubinemia and other comorbid factors (like sepsis and hemolysis) developed kernicterus with TSB ranging from 4 mg/dl to 51 mg/dl.

- Excluding the Collaborative Perinatal Project (CPP) and the studies looking at IQ, of the nine studies primarily looking at behavioral and neurodevelopmental outcomes in patients, only three studies were of high methodological quality. One showed a correlation between bilirubin level and decreased scores on newborn behavioral measurements. One found no difference in prevalence of central nervous system abnormalities at age 4 years when bilirubin was below 20 mg/dl, but infants with bilirubin above 20 mg/dl had a higher prevalence of central nervous system abnormalities. Another that followed infants with bilirubin greater than 16 mg/dl found no relationship between bilirubin and neuro-visual-motor testing at 61 to 82 months of age.
- Six high quality studies (not counting the CPP) showed significant relationship between abnormalities in brainstem auditory evoked potentials and high bilirubin levels. The majority reported resolution with treatment. Three studies reported hearing impairment associated with elevated bilirubin (>16 mg/dl to >20 mg/dl).
- Again excluding CPP, of the eight studies reporting intelligence outcomes in subjects with hyperbilirubinemia, four were considered high quality. These four studies reported no association between IQ and bilirubin level with followup ranging from 6.5 years to 17 years.
- The Collaborative Perinatal Project—with 54,795 live births between 1959 and 1966 from 12 centers in the United States—has, by far, the largest database for the study of hyperbilirubinemia. The study focused only on black and white infants with birthweight $\geq 2,500$ grams, and a comprehensive analysis of the 7-year outcomes of 33,272 subjects was performed. All causes of jaundice were included in the analysis. No consistent association between peak bilirubin level and IQ was found. Rate of sensorineural hearing loss was not related to bilirubin level. Only the frequency of abnormal or suspicious neurologic examination result was associated with bilirubin.
- Short-term studies tend to be of high methodological quality compared to long-term studies but they use tools that have unknown predictive abilities. Long-term studies suffer from high attrition rates of study population and a non-uniform approach to defining “normal neurodevelopmental outcomes.”
- Given the overall diverse conclusions, except in cases of kernicterus with sequelae, the EPC team concluded that the use of a single total serum bilirubin level (within the range described in the studies) to predict long-term behavioral or neurodevelopmental outcomes is inadequate and will lead to conflicting results.

Question 2

What is the evidence for effect modification of the results in Question 1, by gestational age, hemolysis, serum albumin, and other factors?

- The only study that directly addressed the above question used the CPP population and reported that at age 4 years, the frequency of low IQ with increasing bilirubin levels increased more rapidly in infants with infected amniotic fluid. At age 7 years, neurologic abnormalities were also more prevalent in that subgroup of infants.
- In case reports of kernicterus, all four infants with multiple comorbid factors had sequelae.
- Reserve albumin concentration and duration of hyperbilirubinemia will need further studies to understand the nature of neurodevelopment in relation to bilirubin physiology.

Question 3

What are the quantitative estimates of efficacy of treatment at reducing peak bilirubin levels (e.g., number needed to treat at 20 mg/dl to keep TSB from rising)?

- Regardless of different protocols of phototherapy, the NNT for prevention of serum bilirubin level exceeding 20 mg/dl ranged from 6 to 10 in healthy term or near-term infants. This implies that one needs to treat 6 to 10 otherwise healthy jaundiced neonates with TSB ≥ 15 mg/dl by phototherapy in order to prevent the TSB in 1 infant from rising above 20 mg/dl. Phototherapy combined with cessation of breastfeeding and substitution with formula was found to be the most efficient treatment protocol for healthy term or near-term infants with jaundice.

Question 4

What is the efficacy of various strategies for predicting hyperbilirubinemia, including hour-specific bilirubin percentiles?

- For accuracy of various strategies for prediction of neonatal hyperbilirubinemia, 153 articles were included after title and abstract screening. However, only 17 articles were included after full text screening and 10 articles remained after data abstraction—7 percent of the original number. This was the lowest yield of articles among the five questions addressed, suggesting that relatively few prospective or retrospective studies addressed this question without significant design or reporting errors, missing data, or apparent bias.
- A conclusion is difficult to make from these studies. The first challenge is the lack of consistency in defining clinically significant neonatal hyperbilirubinemia. Not only did multiple studies use different levels of total serum

bilirubin to define neonatal hyperbilirubinemia, but the levels of TSB defined as significant also varied by age; age at TSB determination varied by study as well.

The second challenge is the lack of consistency in study populations. These studies were conducted among multiple racial groups in multiple countries, including China, Denmark, India, Israel, Japan, Spain, and the United States. Although infants were defined as healthy term and near-term newborns, these studies included neonates with potential for hemolysis from ABO-incompatible pregnancies, as well as breast-fed and bottle-fed infants. This information was often not specified.

Question 5

What is the accuracy of transcutaneous bilirubin measurements?

- Based on the evidence from the systematic review, transcutaneous measurements of bilirubin by each of the three devices described in the literature—the Minolta Airshields Jaundice Meter™, the Ingram Icterometer, and the SpectRx BiliCheck™—have a linear correlation to total serum bilirubin and may be useful as screening devices to detect clinically significant jaundice and decrease the need of serum bilirubin determinations.
- The Minolta Airshields Jaundice Meter™ appears to perform less well in black infants as compared to white infants, performs best when measurements are made at the sternum, and performs less well when infants have been exposed to phototherapy. This instrument requires daily calibrations and each institution must develop its own correlation curves of TcB to TSB. As a screening test it does not perform consistently across studies as evidenced by the summary ROC curves. The Ingram Icterometer has the added limitation of lacking objectivity of the other methods as it depends on observer visualization of depth of yellow color of the skin.
- The recently introduced BiliCheck™ and Colormate III devices (that utilize reflectance data from multiple wavelengths) appear to be a significant improvement over the older devices, the Ingram Icterometer and the Minolta AirShields bilirubinometer, because of the ability to determine correction factors for the effect of melanin and hemoglobin. In one study, the BiliCheck™ was shown to be as accurate as standard laboratory methods in predicting TSB determined by the reference standard of high performance liquid chromatography (HPLC).

Future Research

- Future research in kernicterus would benefit from a uniform definition of the disease. Making kernicterus a reportable condition coupled with multi-center

cooperation will help to elucidate its epidemiology and other factors in the development of the disease.

- Duration of hyperbilirubinemia, bilirubin binding, and the role of hour-specific bilirubin measurement all merit further investigation.
- Validation of an age-specific (by hour) nomogram for TSB in healthy full-term infants—with evaluation of potential differences by gender, race, and ethnicity, as well as prenatal, natal, and postnatal factors—would be beneficial. Once established, use of the 95th percentile to define clinically significant jaundice would provide uniformity across studies. This would be analogous to the use of age-specific systolic and diastolic blood pressure to define hypertension and age-specific body mass index to define overweight and obesity in children. Validation of a standardized TSB nomogram would either incorporate these potential differences or result in the development of population-specific nomograms if differences were significant. This would be analogous to the use of population-specific growth charts for weight and height percentiles in children.
- Given the interlaboratory variability of measurements of serum bilirubin, future studies should use HPLC as the reference standard along with the routine laboratory methods of TSB in use when evaluating noninvasive measures of bilirubin.
- Validation is needed of new technological advances in the transcutaneous measurement of bilirubin, such as the Bilicheck™ and Colormate III (that have the ability to correct for skin color effects and hemoglobin) in diverse clinical populations. This would address issues that might affect performance such as race, gestational age, age at measurement, phototherapy, sunlight exposure, feeding, and accuracy as screening instruments and for ongoing monitoring of jaundice. Additionally, studies should address cost effectiveness and reproducibility in actual clinical practice.

Ordering Information

The full evidence report from which this summary was taken was prepared for AHRQ by the Tufts New England Medical Center Evidence-based Practice Center, Boston, MA, under contract No. 290-97-0019. It is expected to be available in early 2003. When available, printed copies may be obtained free of charge from the AHRQ Publications Clearinghouse by calling 800-358-9295. Requesters should ask for Evidence Report/Technology Assessment No. 65, *Management of Neonatal Hyperbilirubinemia*. Internet users will be able to access the report online through AHRQ's Web site at www.ahrq.gov.



www.ahrq.gov

AHRQ Pub. No. 03-E005
November 2002
ISSN 1530-440X