

Does Hyperbilirubinemia Damage the Brain of Healthy Full-Term Infants?

Thomas B. Newman, MD, MPH,*
and M. Jeffrey Maisels, MB, BCh†

A full-term, healthy, breast-feeding newborn has a total serum bilirubin of 17 mg/dl. There is no hemolysis. Is this infant at risk for brain damage?

To extract the answer to this question from the literature is a daunting task. The relevant studies span many disciplines—from biochemistry and physiology to hematology, neonatology, neuropathology, and developmental psychology. The authors of these studies have investigated the relationship between a wide assortment of predictor variables (related to bilirubin) and outcome variables (related to brain damage). One author may relate the peak total bilirubin levels to hearing loss, another the duration of hyperbilirubinemia to the intelligence quotient (IQ), and a third the bilirubin binding capacity to abnormalities found on neurologic examination. Thus, synthesizing the results of such studies in a meaningful way is probably impossible. Even when different studies measure approximately the same predictor and outcome variables, the parameters used to measure the effect size may vary greatly. The effect on IQ, for example, may be expressed as mean IQ in different groups of patients, as the proportion of children with IQs below 70, 80, or 90, as a correlation coefficient between bilirubin and IQ, and so on. Finally, the groups of infants studied differ widely, producing discrepant results that are difficult to interpret.

METHODS

These obstacles notwithstanding, we have attempted to make some sense out of the published data. To do this we first limited our review to

* Assistant Professor, Departments of Pediatrics and Epidemiology and Biostatistics, University of California, San Francisco, California; and The Robert Wood Johnson Clinical Scholars Program

† Clinical Professor of Pediatrics, Wayne State University School of Medicine, Detroit; Clinical Professor of Pediatrics, University of Michigan Medical School, Ann Arbor; and Chairman, Department of Pediatrics, William Beaumont Hospital, Royal Oak, Michigan

studies of intact human newborns. Biochemical, cellular, and animal studies have helped to explain how bilirubin might damage the brain and under what clinical circumstances such damage might occur; however, only studies of human neonates can provide an estimate of the magnitude of the effect and the specific clinical circumstances in which it is relevant. We will accept without dispute that it is biologically plausible (indeed likely, under some circumstances) that bilirubin is hazardous to the newborn's brain. Rather than exploring the mechanism, however, we will concentrate on estimating the magnitude of the effect.

Second, we restricted our attention to one predictor variable, or "risk factor": peak total serum bilirubin (TSB). This is the risk factor most commonly measured in studies of the effect of bilirubin on the brain, and it is universally available to those faced with clinical decisions.

Third, we examined only three broad, clinically relevant outcomes: cognitive development (as measured by the IQ), abnormalities on neurologic examination, and hearing. Intermediate outcomes, such as alteration of brainstem-evoked potentials or yellow staining of the brain at postmortem in a baby who died of other causes, will not be discussed, because even if associations with these outcomes exist, their clinical relevance is uncertain.

Fourth, we have reduced the results of different studies to common parameters. For studies of IQ, when possible, we converted other measures of effect size to slopes for a regression of bilirubin on IQ; we thus obtained estimates of expected change in IQ for each mg/dl increase in serum bilirubin. We treated the other two outcomes as dichotomous: definite, or no abnormality on neurologic examination, and the presence or absence of sensorineural hearing loss. For these two outcomes, results are presented as risk ratios, risk differences, absolute risks, and odds ratios per one mg/dl increase in TSB (derived from logistic regression).

Finally, we evaluated only one particularly controversial group of babies: full-term infants without hemolytic disease. Because the effects of hyperbilirubinemia on premature babies and those with hemolytic disease may be different, they must be reviewed separately—a task beyond the scope of the current review. For these two groups of babies, previous reviews may be consulted.^{29,41,47,55}

RESULTS

Effect of Hyperbilirubinemia on Cognition

Studies of the relationship between neonatal hyperbilirubinemia and subsequent cognition are summarized in Table 1. Although the inclusion criteria vary, as do the predictor and outcome variables, these studies include predominantly term babies without hemolysis, and all show little or no effect of bilirubin on IQ or mental development.

The first six studies in the table are based on analyses of various large subsets of the U.S. Collaborative Perinatal Project (CPP). The CPP was a multicenter cohort study of 53,043 women who became pregnant between 1959 to 1965; their infants were followed to age 8 years.⁵ Each infant weighing more than 2500 gm had a total serum bilirubin drawn at age 48 ± 12 hours. If the initial bilirubin concentration was at least

(Text continued on page 337)

Table 1. Effect of Serum Bilirubin Levels on IQ in Infants > 2500 gm

AUTHOR (YEAR)	SUBJECTS	PREDICTOR	OUTCOME	RESULT SUMMARY/COMMENTS
Boggs (1967)	CPP Total N = 20,292 2501-3000 gm N = 4916 ≥ 3001 gm N = 13,568 Includes babies with hemolytic disease	Max TSB categorized in 7 levels	Bayley motor score at 8 months Bayley mental score at 8 months	Birthweight 2501-3000 gm -0.25 pt/mg/dl ≥ 3001 gm -0.16 pt/mg/dl 2501-3000 gm -0.17 pt/mg/dl ≥ 3001 gm -0.09 pt/mg/dl P < 0.005 for all Confounding by race and gestational age is possible P = 0.01 Effect size not given, but can be estimated from figures and P values ~0.06 points/mg/dl P = NS Effect size cannot be estimated Effect size = 0.3 IQ pt/mg/dl based on analysis of Table 1 using logistic regression Analysis may be confounded toward null by race P < 0.001
Scheidt (1977)	CPP Total N = 24,524 This table includes only those > 37 weeks > 2500 gm (N unknown)	Max TSB	Bayley motor at 8 months Bayley mental at 8 months	Effect size† ~-.14 points/mg/dl unadjusted (P = 0.05) ~-.21 points/mg/dl unadjusted (P = NS) P = NS (No trend) P = NS
Naeye (1978)	CPP Total N = 33,064 We excluded 3654 infants with placental evidence of perinatal infection N = 29,410 Includes hemolysis and prematures	Max TSB in 10 categories (highest ≥ 16 mg/dl in 923 babies)	IQ < 90 at 4 years	
Rubin (1979)	CPP subset: N = 366 Selected from 1613 children in Minnesota section of CPP and participants in an educational follow-up study Includes hemolysis and prematures All babies with high TSB and random sample of those with TSB < 10 mg/dl	Max TSB mg/dl < 10 N = 125 11-15.9 N = 164 16-23 N = 77 (14 had ET)	Bayley motor at 8 months Bayley mental at 8 months IQ age 4 years IQ age 7 years	

(Table continued on following page)

1

AUTHOR (YEAR)	SUBJECTS	PREDICTOR	OUTCOME	RESULT SUMMARY/COMMENTS
Broman (1975)	CPP Total N = 24,885 Includes hemolysis and prematures N = 11,497 whites N = 13,388 blacks	Max TSB (as continuous variable) TSB ≥ 20 mg/dl in 266 infants	IQ at age 4 years	Group Whites 0.14 pt/mg/dl Blacks 0.27 pt/mg/dl Effect Size
Broman (1985)	CPP Total N = 34,875 The 994 "Low-achieving" children were matched by IQ with 5964 controls	Max TSB (as continuous variable)	"Low achievement" at age 7 years N = 430 whites N = 504 blacks Lowest achievement subgroup, N = 328 Hyperactive subgroup, N = 150 Poorest readers, N = 103	No association with TSB on multivariate analysis for any of these outcomes "Low achievement" associated with TSB in univariate analysis in blacks only
Bjore (1961)	Total N = 92 Full-term babies Coombs' negative Minor or no complications	Max TSB mg/dl 18-19.9 20-24.9 > 25 Controls "not jaundiced"	IQ < 80 at 2-3 years 1/17 4/42 1/71 5/51	Differences NS
Culley (1970)	Total N = 121 BW > 2500 gm "Non-hemolytic jaundice only" ET to maintain TSB < 20 mg/dl	Max TSB mg/dl Mild or no jaundice 12-16 > 16	IQ < 80 at age 5-6 years 0/48 1/49 0/24	Differences NS

Rosta (1971)	Total N = 102 Babies > 2500 gm with TSB > 15 gm/dl Follow-up on 84 at 8 years	Case series All had TSB > 15 mg/dl	IQ at 8 years	Distribution same as in normal population
Valaes (1980)	Total N = 415 Greek babies enrolled in randomized and nonrandomized trials of prenatal phenobarbital to prevent neonatal jaundice, and re-examined at 5-7 years of age; all were term infants > 2500 gm; based on data for the whole group about 14% had ABO in- compatibility and 5% G6PD deficiency	Comparison of: 182 controls Mean TSB at 4 days 9.5 mg/dl; 6.6% > 16 mg/dl 229 prenatal phenobarbital Mean TSB at 4 days, 4.8 mg/dl; 1.1% TSB > 16 mg/dl Both groups combined: Bilirubin > 16 mg/dl vs. Bilirubin ≤ 16 mg/dl Bilirubin > 16 mg/dl only Highest TSB vs. Lower TSB	Intelligence tests at 5-7 years: Visuomotor test Draw a woman test Georgas vocabulary test	Based on multiple regression Phenobarbital group better $P < 0.025$ Phenobarbital better $P < 0.2$ Phenobarbital better $P < 0.9$
			Same as above	No associations
			Same as above	No associations

(Table continued on following page)

Table 1. *Effect of Serum Bilirubin Levels on IQ in Infants > 2500 gm (Continued)*

AUTHOR (YEAR)	SUBJECTS	PREDICTOR	OUTCOME		RESULT SUMMARY/COMMENTS
Scheidt (1990)	Total N = 1339 Collaborative phototherapy trial: 6-year follow-up of 58% of the 1205 babies who survived 1 year	Phototherapy N = 335 Control N = 360 Max TSB similar in both groups	IQ at age 6 years		Differences NS
Brown (1985)	Includes premature babies (N = 790) and babies with hemolysis		Verbal	Full Scale	Most of difference due to babies < 2000 gm and is probably due to differential loss of follow-up (see text)
			103.6	103.3	
			100.9	101.0	
			100.1		

ABBREVIATIONS: CPP = collaborative perinatal project; HDN = hemolytic disease of the newborn; TSB = total serum bilirubin; prem = premature infant; BW = birthweight; ET = exchange transfusion; GA = gestational age; G6PD = glucose-6-phosphate; and NS = not significant.

* Estimated from regression of TSB on percent with motor score < 27 and percent with mental score < 75 with TSB ranging from 10–15 mg/dl.

† Estimated from the 3 points given in Table 2 in Rubin et al.⁴⁸

10 mg/dl, the measurement was repeated daily until the TSB decreased below 10 mg/dl.^{15,51} The predictor variable in each of the CPP studies is the highest total bilirubin value recorded.

In 1967, in the first such report from the CPP, Boggs et al.⁴ reported a negative association between peak TSB and the Bayley scales of infant development administered at age 8 months. Statistically significant results were obtained for both the motor and mental index in all birthweight groups, although the association became progressively weaker as birthweight increased. These preliminary results were not adjusted for gestational age or race. Further analysis of the same data by Scheidt et al.⁵¹ showed that when the data were stratified by gestational age and race, for babies >2500 gm and >37 weeks' gestation, the effect on mental development was no longer statistically significant, and the effect on motor development was apparently diminished. (Actual estimates of effect size are not provided by Scheidt et al, but the *P* value increased from <0.005 to 0.01 with the same sample size). As before, the effect of bilirubin on both mental and motor development appeared to be greater in the smaller and less mature babies.

These articles caused great concern: Could bilirubin be causing subtle brain damage even at levels previously considered safe? Naeye's³⁶ results were hardly reassuring. He found that the proportion of babies with IQ <90 at age 4 years increased in a dose-response fashion as peak TSB increased. The association between TSB and risk of diminished IQ was strongest in babies with pathologic evidence of placental infection, but was statistically significant in babies without infection as well. (Naeye's study includes low-birthweight and premature infants as well as babies with hemolysis, however, so it probably overestimates the effect size in term, normal birthweight babies without hemolysis.)

Conversely, although Rubin and co-workers⁴⁸ were able to confirm the weak association between 8-month Bayley motor scores and peak bilirubin levels, they found no association with IQ at age 4 or 7, even though premature infants and those with hemolysis were included. They studied only a small subset of the CPP, however: a total of 366 children followed in the Minnesota Educational Follow-up Program. Thus, their failure to observe a significant difference might be due simply to lack of adequate power.⁹

The first monograph by Broman et al.⁶ helps provide the answer.† Broman et al. present correlation coefficients adjusted for race, sex, and socioeconomic index between 169 possible predictors and the Stanford Binet IQ measured at age 4 years in the 24,885 collaborative perinatal study participants for whom bilirubin data and a 4-year IQ test were available. Among the predictors tabulated was the peak total bilirubin level, which had a correlation coefficient with IQ of -0.03 in whites and -0.07 in blacks, both significant at the *P* < 0.01 level. By multiplying the correlation coefficients by the standard deviations of IQ in the two races (16 in whites and 14 in blacks) and dividing by the standard deviation of

† It is puzzling and particularly unfortunate that these and other important data were published only in monographs. As monographs are neither peer reviewed, nor listed in the *Index Medicus*, the information may lie buried for years, known only to a handful of cognoscenti. (Until recently, this information eluded the authors of this chapter.)

maximum bilirubin (3.5 mg/dl³⁰) we can express the effect size in another way: the expected change in IQ for each mg/dl increase in serum bilirubin.⁴⁴ As shown in the table, these estimates are 0.14 IQ points/mg/dl for whites and 0.27 IQ points/mg/dl for blacks.

To compare Broman's results with those of Naeye, we performed logistic regression²⁸ on the tabulated data in Naeye's Table 1, restricting attention to the babies without evidence of infection. The result, which is unadjusted for race, sex, socioeconomic status, or gestational age is an estimated odds ratio of 1.03/mg/dl of bilirubin, which is highly significant ($P < 0.001$). This odds ratio is equivalent to a shift in the mean IQ of approximately 0.3 points/mg/dl increase of TSB⁴—similar to the values obtained by Broman. Because race was not considered, and black infants have lower bilirubin values and lower IQs (than white infants), the true value could be slightly higher. Conversely, the values may be falsely high because of confounding by gestational age or other unmeasured factors (such as intracranial hemorrhage).

In their more recent monograph, Broman and colleagues⁵ studied school achievement rather than IQ. The study design was a case-control study nested within a cohort study.³⁹ Out of the 34,875 children studied at the CPP 7-year follow-up, they identified the 994 subjects who were "low achievers"—either on the basis of standardized tests or failure in school. The whole group of "low achievers" as well as various subgroups listed in Table 1 were compared with 5964 controls matched for IQ, race, and sex. Peak serum bilirubin was not associated with poor achievement overall or in any of the subgroups, with the exception of the univariate analysis of the lowest-achievement subgroups in blacks only.

Several smaller studies, all from Europe, have also examined the effect of bilirubin on IQ in term babies without hemolysis. These are summarized in Table 1. Bjure et al.,³ Culley et al.,¹¹ and Rosta et al.⁴⁶ all found no apparent effect of increasing bilirubin levels on IQ. The sample sizes of all of these studies were too small to rule out a small or moderate-sized effect, however.

Two randomized trials are relevant to this issue. The first, by Valaes et al.,⁵⁶ examined the effect of prenatal maternal phenobarbital administration on neonatal bilirubin levels and subsequent development. In the first phase of the study, women were randomly assigned to receive phenobarbital or a placebo. In the second phase, group assignment was based on convenience: Women who were able to attend a late prenatal visit were placed on phenobarbital. The authors noted that the women assigned to phenobarbital in the second phase were more likely to live in

⁴ This estimate is based on a logistic regression analysis on the third column of Naeye's Table 1 (percent with IQ < 90 at age 4 years in those with no infection). The values for bilirubin used in the model were obtained by taking approximate midpoints of the 10 intervals down the left side of the table. Values we used were: 3, 5, 6, 7, 8, 9, 11, 13, 15, and 17. The resulting logistic model was that the odds of IQ < 90 were $\exp(-1.0445 + 0.0303 * \text{TSB [in mg/dl]})$. The probability of abnormality can be obtained from odds by $P = \text{odds} / (1 + \text{odds})$. To convert probability of IQ < 90 obtained from this logistic model to the estimated difference in mean IQ, we assumed that IQ was normally distributed with standard deviation of 16 points.⁶ We then used tables of the normal distribution to determine how large a difference in means of the two groups would be required to obtain the predicted difference in expected proportions < 90.

cities and to be of higher social class but attempted to compensate for this with multiple regression.

The infants of mothers who received phenobarbital performed better on three tests of development; of these only the difference in the "visuomotor integration test" was statistically significant ($P < 0.025$). The authors also combined the two groups and simply compared peak serum bilirubin level with test performance. In none of these combined analyses was serum bilirubin a significant predictor of cognitive performance, even at high levels. (A significant effect on hearing occurred, which will be discussed later.)

How could phenobarbital be associated with improved performance when lower bilirubin levels were not? If better performance was the result of phenobarbital's ability to decrease serum bilirubin, we would expect the relationship between bilirubin and outcome to be *stronger*, not weaker, than the association between phenobarbital and outcome. One explanation for the finding is that the association is due to chance—results were significant for only one of several outcomes tested. Another is that TSB is a poor marker for the real risk factor (free bilirubin, perhaps), which phenobarbital successfully reduces. Finally, it is possible that, despite the use of a multiple regression model, the effect is due to confounding by social factors associated with the nonrandom assignment to the phenobarbital group in the second phase of the study.

In the National Institute of Child Health and Human Development (NICHD) collaborative phototherapy trial^{38,50} infants > 2500 gm were randomly assigned to receive or not to receive phototherapy when the serum bilirubin level reached 13 mg/dl. Because the predictor variable in this study is phototherapy (versus no phototherapy) it provides only indirect (though clinically relevant) evidence on the question of whether bilirubin affects cognition. In fact, because in babies > 2500 gm phototherapy was not begun until the serum bilirubin reached an average of 15.7 mg/dl the *peak* serum bilirubin levels in the two groups were practically identical.⁸ A difference existed in the duration of hyperbilirubinemia, which can be expressed as a difference in areas under the two curves of serum bilirubin plotted against age. This difference was about 12.5 (mg/dl)/(days) in babies > 2500 gm. In the babies weighing 2000 gm — 2499 gm and those < 2000 gm the differences in peak bilirubin levels were about 1 mg/dl and 6 mg/dl, respectively, and the areas under the curves were substantially greater in the control infants. These differences in hyperbilirubinemia were not associated with a difference in mean IQ. Although a modest trend occurred (2–2.5 points) slightly favoring the phototherapy group, the numbers tabulated include 790 babies < 2000 gm (66% of original cohort). In this group a differential loss to follow-up occurred (i.e., a significant excess of black infants lost from the phototherapy group) that might explain a difference in IQ scores.

Effect of Hyperbilirubinemia on Neurologic Abnormality

Studies of serum bilirubin as a risk factor for neurologic abnormality are summarized in Table 2. As before, we'll begin with reports from the CPP, followed by the smaller observational studies and two experiments.

(Text continued on page 344)

Table 2. Neurological Abnormalities in Infants > 2500 gm Without Hemolytic Disease

AUTHOR (YEAR)	SUBJECTS	PREDICTOR	OUTCOME	RESULT SUMMARY/COMMENTS
		Max TSB mg/dl	Def. (N)	Abnl. (%)
		N	(N)	(%)
		0-9	2585	1.4
		10-14.9	300	1.3
		15-19.9	75	1.3
		≥20	35	0
		Total	2995	40 (1.3%)
				(Black and white combined)
				Neurologic Examination
				Suspect or Abnl.
				(%)
				12
				12
				21
				20
				374 (12%)
Hardy (1971)	CPP participants followed at Johns Hopkins Total N = 2995 > 2500 gm Blacks N = 2370 Whites N = 625			OR (for defined abnormality) = 0.975 mg/dl TSB (P > 0.3) 95% CI = (0.87 - 1.10) OR (for suspected abnormality) = 1.03 mg/dl TSB (P = 0.05) 95% CI = (1.00 - 1.07)
Naeye (1978)	CPP Total N = 39,468 We excluded 4946 infants with placental infection Includes babies with hemolysis and premature babies	Max TSB 10 categories mg/dl		Abnormal neurologic evaluation at 7 years
		0-11.9	(a) 4243/35926	= 11.8%
		12-15.9	(b) 308/2253	= 13.7%
		≥16	(c) 215/1289	= 16.7%
				OR = 1.016 mg/dl increase in TSB (P < 0.001) 95% CI = (1.008 - 1.024)
			(a) vs (b)	RR 1.16 Δ = 1.9%
			(a) vs (c)	RR = 1.4 Δ = 4.9%
				No control for confounding
Mores (1959)	Total N = 54 Full-term infants with TSB > 20 mg/dl	Case series Max TSB mg/dl		Abnormal examination for "psychosomatic development"
		20-25	0/35	
		25-30	0/10	
				0/54 abnormal

Table 2. Neurological Abnormalities in Infants > 2500 gm Without Hemolytic Disease (Continued)

AUTHOR (YEAR)	SUBJECTS	PREDICTOR	OUTCOME	RESULT SUMMARY/COMMENTS
Holmes (1968)	Total N = 195 Babies > 2500 gm No RDS infection, congenital malformations ET at 20 mg/dl Follow-up on 80 infants	Max TSB mg/dl Control (no jaundice 6-12.5 > 15	Abnormal neurologic examina- tion at 4.5-7.5 years 0/17 0/29 0/34	RR Undefined $\Delta = 0$ (vs control) $\Delta = 0$ (vs control)
Culley (1970)	Total N = 121 BW > 2500 gm Nonhemolytic jaundice	Max TSB mg/dl No or mild jaundice 12-16 > 16	CNS deficit at 5-6 years 0/48 1/49 (minimal clumsiness, IQ = 85) 1/24 (36-week premature infant with spastic diplegia, IQ = 93)	RR = ∞ $\Delta = 2\%$ $\Delta = 4.2\%$
Bengtsson (1974)	Total N = 226 Healthy, term infants all with negative Coombs' tests	Max TSB mg/dl < 20 > 20 ABO compatible ABO incompati- ble	Abnormal neurologic examina- tion at 6.5-13 years 1/115 = 0.87% 2/111 = 1.8% 1/66 1/45 Abnormal neurologic examina- tion at 1 year (Cerebral palsy, hypotonia, or abnormal movement)	Difference NS RR = 2.1 $\Delta = .93\%$ Low power for rare outcomes Examiners blinded

Scheidt (1990)	Total N = 1339 Collaborative phototherapy trial Follow-up on 63% of 415 babies > 2000 gm who survived 1 year All exchanged at 20 mg/dl	Phototherapy Control	1/121 0/117	
Valaes (1980)	Total N = 415 Greek babies enrolled in randomized and nonrandomized trials of prenatal pheno- barbital to prevent neonatal jaundice; based on data for the whole group about 14% had ABO incompatibility and 5% G6PD deficiency	182 controls Mean TSB at 4 days 9.5 mg/dl; 6.6% > 16 mg/dl 229 prenatal Phenobarbital Mean TSB at 4 days 4.8 mg/dl; 1.1% TSB > 16 mg/dl	1/182	Abnormal neurologic examina- tion at 5-7 years No difference between phenobarbital and control No relation to degree of jaundice
			1/229	
				Coordination score

ABBREVIATIONS: CPP = collaborative perinatal project; TSB = total serum bilirubin; BW = birthweight; ET = exchange transfusion; CA = gesta-
tional age; OR = odds ratio; RR = risk ratio; Δ = risk difference; max = maximal; CNS = central nervous system; and NS = not significant.

Hardy and Peebles¹⁵ studied the Johns Hopkins cohort from the CPP, and found no relation between serum bilirubin levels and the risk of definite neurologic abnormality at 1 year in babies > 2500 gm. Although the study included nearly 3000 babies, relatively few babies were at the highest levels of serum bilirubin. To estimate how strong a relationship might have been missed, we performed a logistic regression analysis of the tabulated data.^{||} As expected, the estimated odds ratio was close to 1, indicating no association. Of note is that the upper limit of the 95% confidence interval reaches only 1.10 per mg/dl of bilirubin. (This is equivalent to $1.10^5 = 1.6$ per 5 mg/dl of bilirubin.) Thus although too small to rule out a small to moderate effect, the Hardy study provides evidence *against* a strong relationship between serum bilirubin and definite neurologic abnormality.

When questionable or suspect abnormalities were included, however, a (barely) statistically significant association occurred. Again, we quantified the association using logistic regression. This time the estimated odds ratio was 1.03 mg/dl of bilirubin, with 95% confidence interval from 1.00 to 1.07. Because the Hardy study excluded babies based on birthweight rather than gestational age, some of this effect may be due to prematurity. In addition, no attempt was made to exclude babies with hemolysis. Finally, we do not know whether the examiners were blinded to the children's past bilirubin levels, a crucial issue given the "soft" neurologic findings for which the results were significant.

Naeye³⁶ reported results in substantial agreement with those of Hardy; using Naeye's data, we calculated an odds ratio of 1.016 per mg/dl serum bilirubin.^{||} As mentioned previously, these data are somewhat difficult to interpret because low-birthweight and premature babies as well as infants with hemolysis were included, although we were able to exclude babies with placental evidence of infection. The outcome variable ("abnormal neurologic evaluation") in Naeye's study must have included suspected abnormalities, because the overall rate of about 12% was similar to that found by Hardy and Peebles for suspected plus definite abnormalities. Again, no mention was made of blinding in either study.

Mores et al.,³⁵ Killander et al.,²³ Fohl and Lombos,¹⁴ Holmes et al.,¹⁷ Culley et al.,¹¹ and Bengtsson and Verneholt² have reported smaller studies of the risk of neurologic abnormality. All of these studies found

^{||} For our logistic regression, we combined entries for whites and blacks in Table 5 from Hardy and Peebles's article.¹⁵ For bilirubin values we used approximate midpoints of the table entries: 6, 12, 17, and 22 mg/dl. The resulting logistic models were:

$$\text{Odds of definite neurologic abnormality} = \exp(-4.1285 - 0.0250 * \text{SB [in mg/dl]})$$

$$\text{Odds of definite or suspect abnormality} = \exp(-2.2055 + 0.036 * \text{serum bilirubin [in mg/dl]})$$

^{*} This estimate is based on a logistic regression analysis on the right-most column of Naeye's Table 1 (percent with neurologic abnormalities at age 7 years in those with no infection). The values for bilirubin used in the model were obtained by taking approximate midpoints of the 10 intervals down the left side of the table. Values we used were: 3, 5, 6, 7, 8, 9, 11, 13, 15, and 17. The resulting logistic model was that the odds of abnormality were $\exp(-2.0851 + 0.0159 * \text{TSB [in mg/dl]})$.

low rates of definite neurologic abnormalities, but all had sample sizes too small to be totally reassuring. If the results of these studies are combined with those of Hardy and Peeples, however, abnormal neurologic examinations were found in 7 of 462 infants (1.5%) with TSB of at least 15 mg/dl and in 5 of 323 infants (1.5%) with TSB of at least 20 mg/dl. These rates, although obtained by different examiners of different infant populations, are similar to the $35/2585 = 1.4\%$ reported by Hardy for babies with maximum TSB of less than 10 mg/dl. Moreover, they probably overestimate the risk of TSB < 30 mg/dl, because four of the six abnormal infants had TSB > 30 mg/dl.^{2,14} Thus these smaller studies, taken together, suggest that serum bilirubin has little effect on the risk of neurologic abnormality.

Finally, the results of neurologic examinations in the two randomized trials are worth discussing. In the trial of prenatal phenobarbital, 1 of 229 infants who received phenobarbital was neurologically abnormal at 6 years, compared with 1 of 182 controls. (Neither of the two neurologically abnormal children appears to have been significantly jaundiced; one was reported to have had "mild jaundice," the other an "unremarkable perinatal history.") In the NICHD collaborative phototherapy trial, the neurologic examination at age 1 year was abnormal in 1 of 121 infants in the phototherapy group, compared with none of 117 infants in the control group. Because of the rarity of the outcome, however, neither of these trials (nor the combination of both) has sufficient power to rule out a beneficial effect of lowering serum bilirubin; however, both suggest that the benefit will be small. Both studies are also consistent with the results of the CPP and smaller studies cited earlier, suggesting that elevated bilirubin is not a risk factor for neurologic abnormality in this population.

Effect of Hyperbilirubinemia on Hearing

Sensorineural hearing loss is a common manifestation of kernicterus,²⁹ and hyperbilirubinemia has a transient effect on brainstem auditory-evoked responses, even in apparently healthy term babies.^{37,43,57} This suggests that the auditory pathways might be particularly sensitive to hyperbilirubinemia, even in babies without hemolytic disease.

The relevant studies are summarized in Table 3. One study not previously mentioned (nor referenced in any other review of this subject that we have seen) is by Lassman et al.²⁶ This monograph summarized the results of analyses correlating 901 possible predictors measured in the CPP with indices of speech, language, and hearing created from 306 measurements on the 3- and 8-year examinations. Sensorineural hearing loss was defined as a loss of at least 30 decibels (dB) at two or more of the four test frequencies, with an average airborne gap across the same frequencies of 10 dB or fewer. Peak SB, measured as previously described, had *no effect* on any speech, language, or hearing test performance. In fact, although the specific data are not tabulated, it is noted in the text that the incidence of sensorineural hearing loss was highest in the babies with the *lowest* bilirubin levels, as was the incidence of conductive hearing loss.

(Text continued on page 348)

Table 3. *Effect of Serum Bilirubin Levels on Hearing in Infants > 2500 gm*

AUTHOR (YEAR)	SUBJECTS	PREDICTOR	OUTCOME	RESULT SUMMARY/COMMENTS
Lassman (1980)	CPP Statements about hearing loss apparently refer to the approximately 14,900 babies who had complete speech, language, and hearing evaluations at ages 3 and 8 years Includes premature babies and babies with hemolysis	Max TSB	Sensorineural hearing loss at 3 and 8 years Speech and language evaluations at 3 and 8 years	No effect of bilirubin on speech language, or hearing. "Actually, the highest incidence of sensorineural hearing loss occurred among babies whose bilirubin concentration never rose above 5 mg/dl."
Bjure (1961)	Total N = 92 Full-term babies Coombs' negative Minor or no complications	Max TSB mg/dl 18-19.9 20-24.9 ≥ 25 Controls	Hearing loss 0/14 0/34 0/5 0/39	No hearing loss found
Holmes (1968)	Total N = 195 Babies > 2500 gm No RDS, infection, or congenital malformations ET at 20 mg/dl 195 eligible	Max TSB mg/dl Control 12.5 > 15	Hearing impairment 0/17 0/29 0/34	No hearing impairment found

Culley (1970)	Total N = 121 BW > 2500 gm Nonhemolytic jaundice	No or mild jaundice 12-16 ≥ 16	Normal or (if abnormal) owing to catarrhal conditions in all, but denominator not given	No hearing impairment found
Bengtsson (1974)	Total N = 226 Healthy, term infants all with negative Coombs' tests	Max TSB mg/dl < 20 ≥ 20 ABO compatible ABO incompati- ble	Hearing loss 0/115 4/111 1/66 3/45	Trend toward ↓ hearing with ↑ TSB, especially in ABO incompatible group but not statistically significant
Valaes (1980)	Total N = 415 Greek babies enrolled in randomized and nonrandomized trials of prenatal pheno- barbital to prevent neonatal jaundice, and re-examined at 5-7 yrs of age; all were term singleton infants > 2500 gm; based on data for the whole group about 14% had ABO incompatibility and 5% G6PD deficiency	Phenobarbital group control group Max TSB mg/dl (a) No jaundice (b) < 12 (c) 12-16 (d) 16 < 12 ≥ 12	Sensorineural hearing loss 4/233 = 1.7% 6/182 = 3.3% 2/243 = 0.82% 2/107 = 1.9% 3/32 = 9.4% 3/33 = 9.1% 4/350 = 1.1% 6/65 = 9.2%	Examiners blinded RR = 1.9 Severity of sensorineural hearing loss not given Δ = 1.6% P = NA (a) vs (b) P = NS (a) vs (c) P < 0.001 (a) vs (d) P < 0.001 RR = 8.1 Δ = 8.1% P < 0.001

ABBREVIATIONS: CPP = collaborative perinatal project; TSB = total serum bilirubin; BW = birthweight; ET = exchange transfusion; CA = gesta-
tional age; Max = maximal; RR = risk ratio; Δ = risk difference; and NA = not applicable.

The studies by Bjure et al.,³ Holmes et al.,¹⁷ and Culley et al.¹¹ all report no sensorineural hearing loss in any of the babies tested, but all of the sample sizes were small. Bengtsson et al.² found a trend toward a higher rate of hearing loss (4 of 111 with TSB more than 20 mg/dl versus none of 115 with TSB < 20 mg/dl), but it was not statistically significant, perhaps because of small sample size. Although three of the four babies in Bengtsson's study had ABO incompatibility, all had negative Coombs' tests, making it difficult to attribute their hearing loss of hemolysis.

Valaes⁵⁶ did report a statistically significant relationship between increasing serum bilirubin levels and the risk of sensorineural hearing loss in the prenatal phenobarbital trial. Interpretation of this result is complicated by the lack of information on the severity of the hearing loss, how it was measured, whether those measuring it were blinded to bilirubin values, and the inclusion of babies with hemolytic disease. Although most of the subjects did not have hemolytic disease, of the 45 infants with bilirubin values > 16 mg/dl, 13 had ABO incompatibility, 7 were premature, and 1 was glucose-6-phosphate dehydrogenase (G6PD) deficient, perhaps accounting for at least a few of the six children with sensorineural hearing loss in the high bilirubin groups.

DISCUSSION

Selection of Studies: Confounding and Effect Modification

Although we excluded all studies that dealt mainly or exclusively with premature infants or those with hemolytic disease of the newborn, some of the studies reviewed (notably those from the CPP) did include some premature babies or infants with hemolysis. To understand the rationale behind our selection process, and how it might affect our conclusions, it is important to understand two key epidemiologic concepts: confounding and effect modification.

In the present context, both of these terms refer to distortion or modification (by extraneous factors) of an observed relationship between bilirubin and outcome. For example, compared with term infants, premature babies tend to have higher bilirubin levels and also a higher risk of adverse neurologic outcome. Thus, if we find that babies with high bilirubin levels do worse, the first question we must ask is whether it is because they are less mature. Unless we examine the association between bilirubin and outcome separately at each gestational age, some or all of the differences we attribute to bilirubin may be due to differences in maturity.

In Table 4, A and B provide a numeric example in which the entire association between bilirubin and neurologic abnormality is due to confounding by prematurity. The true relationship between bilirubin and outcome is the same in both premature and term babies, can be summarized with one number (in this instance, a risk ratio of 1), and can be seen by stratifying the data by prematurity.

Effect modification, illustrated in Table 4, A and C, is similar to confounding with one essential difference. Effect modification means that the effect of bilirubin on outcome is *modified* by prematurity. Once

Table 4. *Hypothetic Examples of Confounding and Effect Modification Distorting Relationship Between Hyperbilirubinemia and Neurologic Abnormality**

A. COMBINED RESULTS ON PRETERM AND TERM INFANTS				
	Abnl	N	% Abnl	
Bilirubin ≥ 20	30	300	10	RR = 10/6.7 = 1.5
Bilirubin < 20	60	900	6.7	
B. STRATIFIED RESULTS DEMONSTRATING CONFOUNDING				
1. Preterm babies				
	Abnl	N	% Abnl	
Bilirubin ≥ 20	20	100	20	RR = 20/20 = 1
Bilirubin < 20	20	100	20	
2. Term babies				
	Abnl	N	% Abnl	
Bilirubin ≥ 20	10	200	5	RR = 5/5 = 1
Bilirubin < 20	40	800	5	
C. STRATIFIED RESULTS DEMONSTRATING EFFECT MODIFICATION				
1. Preterm babies				
	Abnl	N	% Abnl	
Bilirubin ≥ 20	18	100	18	RR = 18/12 = 1.5
Bilirubin < 20	12	100	12	
2. Term babies				
	Abnl	N	% Abnl	
Bilirubin ≥ 20	12	200	6	RR = 6/6 = 1
Bilirubin < 20	48	800	6	

ABBREVIATIONS: Abnl = abnormal neurologic examination; and RR = risk ratio.

* In the combined table (panel A), neither confounding nor effect modification can be discerned. The results in panel A could be distorted by confounding, as illustrated in panel B or by effect modification as in panel C.

again, the true effect of bilirubin on outcome is disclosed by stratification. With effect modification, however, the relationship differs in different strata. Thus when effect modification occurs, any statement about the effect of bilirubin on outcome must be accompanied by a statement about which gestational age is being discussed—no one number can summarize the effect.

The possibility of confounding and effect modification has important implications. First, studies seeking to demonstrate a causal relationship between bilirubin and outcome must also measure and control for other determinants of outcome associated with bilirubin levels. Second, because effect modification by hemolytic disease or prematurity is likely, the relationship between bilirubin and outcome must be examined separately in babies with and without hemolytic disease, and in preterm and term babies. The risk relationships may be different in these groups.

In most studies on the effect of hyperbilirubinemia, one or more probable confounders or effect modifiers has not been considered. Many

of these studies can nonetheless provide useful information, because the direction of the potential error can be predicted. A confounder such as prematurity or intracranial hemorrhage, which is positively associated with both bilirubin and adverse neurologic outcome can only lead to distortion of the effect estimate *away* from the null—to a false-positive but not a false-negative result. Thus, studies in which these two possible confounders are not considered are still valid if they have negative results. Similarly, hemolytic disease, a probable effect modifier, probably *increases* the risk associated with hyperbilirubinemia. Therefore, if our interest is in babies with no hemolysis, studies that include babies with hemolytic disease are still of interest if they show little effect, because they provide an upper limit for the effect size. Conversely, if such studies do show an effect, the upper-limit estimate they provide will be too high to be informative about babies without hemolysis.

To illustrate this point, a few studies we did not tabulate are worth mentioning briefly. First, in the classical studies by Allen et al.,¹ Hsia et al.,¹⁸ and Mollison et al.³²⁻³⁴ of babies with erythroblastosis, the highest bilirubin levels were associated with an incidence of kernicterus of 30% to 50%. Interestingly, it is from these studies of babies with severe Rh hemolytic disease that the tradition of keeping the bilirubin below 20 mg/dl was derived; yet that rule has also been applied to babies with no apparent hemolysis, despite strong evidence (summarized earlier) that such babies are at much lower (if any) risk of bilirubin-induced neurologic damage.

Similarly Hyman et al.,²⁰ in a careful follow-up study of 405 infants with hemolytic disease of the newborn compared 119 children whose maximum indirect serum bilirubin exceeded 20 mg/dl with 286 children with lower bilirubin levels. Children with higher bilirubin levels had 3 times the risk of neurologic abnormality (20% versus 6.6%), and 4.4 times the risk of sensorineural hearing loss (9.2% versus 2.1%).

Because effect modification by hemolysis (or some other factor related to erythroblastosis) is likely, the preceding studies are only marginally relevant to (those dealing with) babies without hemolysis. Moreover, these studies may be only marginally relevant to babies *with* hemolysis born in 1990, or later because many other potential effect modifiers present 20 to 40 years ago are no longer operative. In the early erythroblastosis studies, for example, artificial rupture of membranes was routine, leading to premature delivery.¹⁸ These infants were therefore at much higher risk for kernicterus. Thus the strong relationship between elevated serum bilirubin and kernicterus observed in these studies may be due partly to effect modification by then-current obstetric and neonatal management practices, long since abandoned.

Another example of a possible effect-modifying practice is provided by Hyman et al.¹⁸ In that study all babies receiving exchange transfusions (and many others as well) received prophylactic streptomycin, an ototoxic antibiotic. Although babies with elevated bilirubin levels had a higher risk of hearing loss even when those not receiving streptomycin were excluded from the comparison group (thus excluding confounding), it is impossible to exclude effect modification because all infants with high bilirubin levels received streptomycin. It is certainly possible

that a high serum bilirubin level is more ototoxic when accompanied by streptomycin.

In summary, because confounding and effect modification by many factors will predictably *increase* the observed effect size in the low-risk group (well, term babies without hemolysis) studies in which such factors are not considered provide useful information only if little or no effect is found. (Conversely, if one is interested in a high-risk group—e.g., sick premature infants with Rh disease—studies that include babies at low risk will tend to give a *lower* limit for the effect size and will therefore be most informative when they find large effects.)

Are We Measuring the Right "Risk Factor"?

Two additional studies are worth discussing. Odell et al.⁴² examined 32 children at 4 to 7 years of age in whom bilirubin levels were measured in the neonatal period. (We did not tabulate this study because 16 of the 18 abnormal infants were either premature or had hemolytic disease.) They found that TSB was unrelated to abnormal performance, but that the bilirubin saturation index (a measure of the saturation of serum albumin binding sites by bilirubin) was strongly related to subsequent abnormality. The effect size in this study was incredible: a risk difference of 58% (i.e., "brain damage" occurred in 76% of those with high saturation, compared with 18% of those with low saturation). The small sample size, however, could lead to a gross overestimation of the effect simply by chance. Furthermore, two thirds of the infants enrolled were lost to follow-up, raising the possibility of substantial bias. Finally, an effect size this enormous is almost impossible to miss, yet no later studies have confirmed these results.

Johnson and Boggs²⁵ reported results of detailed neurologic and psychometric examinations on 83 infants with a history of neonatal jaundice. (Again, we did not tabulate the results because 44 [53%] of the infants had hemolytic disease, and 27 [33%] were premature.) Although (unlike Odell et al.) they did find a relationship between peak indirect bilirubin > 15 mg/dl and risk of abnormality, the strongest relationship was with duration of hyperbilirubinemia; the longer the hyperbilirubinemia lasted, the greater the risk.

These studies suggest that one possible explanation for the poor correlation between maximum serum bilirubin and outcome is that maximum SB is only weakly associated with the "true" risk factor—be it free serum bilirubin or duration of hyperbilirubinemia. As mentioned previously, this intriguing possibility could also explain the otherwise puzzling results of Valaes et al.,⁵⁶ who found phenobarbital exposure to be a stronger predictor of performance on a visuomotor integration test than was serum bilirubin. (This hypothesis is also supported by some biochemical and animal evidence not reviewed here.^{10,29,47})

What About Measurement Error? An important barrier to deciding whether bilirubin causes brain damage is the relative inaccuracy of bilirubin measurements. This subject is discussed in detail elsewhere in this issue. When different laboratories measure total bilirubin levels on standard samples (with identical bilirubin concentration) the coefficient of variation (the standard deviation divided by the mean) is about 10%

when the total bilirubin is in the range of 5 to 20 mg/dl.^{42,52,59} Coefficients of variation within the same laboratory (i.e., on duplicate analyses of the same specimen) are about half as large.⁵⁹ Thus, if the true serum bilirubin is 20 mg/dl, a 10% coefficient of variation between laboratories means that about 68% of measurements of bilirubin in different laboratories will be between 18 and 22 mg/dl, and 95% will be between 16 and 24 mg/dl. Within the same laboratory, 68% of measurements will fall between 19 and 21 mg/dl, and 95% between 18 and 22 mg/dl.

What effect will this random error have on studies of the relationship between measurements of bilirubin and outcome? Interlaboratory variability will interfere primarily with the interpretation of descriptive studies. For example, the distribution of bilirubin levels and the frequency of abnormalities at a given level will vary depending on whether the laboratory's bilirubin levels run high or low. Analytic studies, conversely, in which babies with high bilirubin levels are compared with babies (from the same institution) with lower levels will be less affected by interlaboratory variation, because to some extent the errors will cancel out each other. Such studies will still be affected by intralaboratory variation (random error), however, which in general will reduce the apparent effect of bilirubin on outcome, and make the effect more difficult to detect statistically.

Measurements of the outcome variable are also subject to error. The IQ is an imperfect measurement of cognition, the physical examination an imperfect measurement of neurologic damage, and so on. Random and nondifferential errors in these measurements will also make a true effect of bilirubin more difficult to detect. Differential errors in outcome measurements, conversely, can lead to a falsely positive result.¹⁹ Such a differential measurement error could occur if an examiner, consciously or unconsciously, was more likely to identify a borderline neurologic abnormality in a patient known to have been jaundiced. Studies most susceptible to differential measurement errors are those in which the outcome measure is not objective (e.g., "suspect" neurologic examination), and the observer is not blinded to the bilirubin status of the subject.

In summary, with the exception of unblinded studies of "soft" outcomes, the primary effect of measurement error will be to bias the results toward the null. Does this mean that null findings can be dismissed as possibly due to measurement error? We think not. First, the measurement error for bilirubin within a laboratory is not excessive compared with other clinical measurements.⁴⁵ Although standardization of measurements in different laboratories has been problematic, in multivariate analyses of the CPP, indicator variables for site of birth were included in the regression model,^{6,26} thus allowing bilirubin to be related to outcome separately for each site. More important, to the extent that research studies reflect measurement errors present in the outside world, they give results that are more, not less, useful clinically. A clinician caring for a jaundiced baby needs to know how well routine bilirubin levels from his laboratory predict outcome. A possible relationship between "true" bilirubin (but not measured bilirubin) and neurologic outcome is of little relevance clinically, because the clinician has no way of knowing what the "true" bilirubin is.

Similar considerations apply to errors in measurement of the outcome. When a study finds bilirubin is not associated with IQ or neurologic abnormalities, it is always possible to argue that our measures of cognition and neurologic handicap are simply too crude to identify subtle effects. Such arguments have little merit. Rather than worrying about adverse effects so subtle that they cannot be identified, or are only detectable statistically in groups of tens of thousands of patients, we should redirect our efforts to more important risk factors (such as socioeconomic status and birthweight) that have large, easily measured effects on the same outcomes.

Clinical Versus Statistical Significance. Most of the smaller studies tabulated earlier report no association between serum bilirubin and neurologic outcome. Conversely, the CPP—the largest, most definitive study, has found at least a few statistically significant relationships.

Despite the *statistically* significant results in the studies just mentioned, however, the CPP provides excellent evidence that serum bilirubin is not associated with subsequent IQ in any *clinically* important way. Correlation coefficients of -0.03 and -0.07 mean that even if we assume that the relationship is entirely causal, bilirubin explains only 0.09% and 0.49% of the variance in IQ in whites and blacks, respectively. These correlations are statistically significant only by virtue of the enormous sample size. Expressed another way, the effect size for IQ was about 0.15 to 0.3 IQ points/mg/dl of serum bilirubin (thus the *worst* case estimate for the risk of a serum bilirubin of 20 mg/dl versus 10 mg/dl is a loss of 3 IQ points). For suspect neurologic examinations, the odds ratio is about 1.02/mg/dl of bilirubin. For hearing, given the trend in the opposite direction in the CPP, it seems unlikely that any adverse effect occurs at all.

To put these effect sizes into perspective, we should consider the magnitude of change in bilirubin presently achievable by various interventions. Exchange transfusion, too hazardous a procedure for use in low-risk babies, results in a drop (after the rebound) of about 30% to 40% of the peak bilirubin level.²⁹ Phenobarbital given prenatally lowered the mean serum bilirubin at 4 days by about 5 mg/dl.⁵⁶ Phototherapy, when used therapeutically (rather than prophylactically), as it was in the collaborative phototherapy trial, has little effect on peak serum bilirubin in term infants. Even when used prophylactically in premature babies, the effect on mean peak serum bilirubin was only a few mg/dl.⁸ Tin protoporphyrin reduced the mean serum bilirubin by only 2 to 3 mg/dl in a group of ABO incompatible babies.²²

Thus, even if the small observed associations were completely causal, and even if interventions to reduce hyperbilirubinemia could completely eliminate the risk, the potential gain in IQ would be only about 0.75 to 1.5 points—an effect only about 5% to 10% as large as the effect of race or socioeconomic status.^{5,6} Similarly the predicted reduction in definite or suspected neurologic abnormality is also small; decreasing the serum bilirubin from 18 to 13 mg/dl, for example, would produce a predicted change in the incidence of definite or suspected abnormality from 14% to 13%. Because the effect size is so small, it is most unlikely that a search for new interventions directed at preventing

or treating hyperbilirubinemia will be rewarding, particularly in term babies without hemolysis. So little is to be gained that the therapy would be ethically acceptable only if its risk were virtually zero, and studies to demonstrate such a level of safety would require enormous sample sizes. Surely our time and resources would be much better spent attempting to narrow the gaps caused by race, socioeconomic status, or lack of prenatal care.

LIMITATIONS AND IMPLICATIONS

To Whom Do These Results Apply?

We have found, in term, well babies without hemolysis, remarkably little evidence of adverse effects of bilirubin on IQ, neurologic examination, or hearing. In fact, the results from the CPP provide data to the contrary that are reassuring. This conclusion is in agreement with others who have specifically considered this group of babies.^{16,24,54,58}

Some difficulties exist, however. We do not always know which babies have hemolysis, and different investigators have used different criteria that, in many cases, were not explicit. Are babies with ABO incompatibility but a negative Coombs' test at increased risk? What about babies with a positive Coombs' test, but a normal hemoglobin and reticulocyte count? Are really good data available to support the belief that babies with hemolytic disease (treated today) are at greater risk for brain damage than those without hemolytic disease?

Asian or Mediterranean babies at risk for G6PD deficiency present special problems.^{7,12,13,27,49,53} Because few such babies were in the CPP, and they were not analyzed separately, it is likely that an adverse effect would have been missed. Unfortunately, the smaller studies in Tables 1 to 3 are mostly from Scandinavia and Hungary, where G6PD deficiency is uncommon. Thus for babies at risk of G6PD deficiency, even if the blood types are compatible and the direct Coombs' test is negative, we do not know whether their jaundice is benign.

Similar considerations apply to the determination that the baby is "term" and "well." In many studies, the definition of "term" was based on a birthweight > 2500 gm; in others it was based on estimated gestational age. Do our conclusions apply to 2400-gm term babies or to 2700-gm premature infants? Does the adjective "well" exclude a vigorous infant born after 32 hours of ruptured membranes, with a moderately elevated white blood cell count?

The answers to these questions are unknown, and clinicians must do their best to extrapolate from existing studies. Many babies, however, do fit unambiguously into a low-risk group, including the baby with a TSB of 17 mg/dl described at the beginning of this review. How should we manage such an infant?

Based on our review, it would appear that the biggest risk to our hypothetical infant is that he will be subjected to an exchange transfusion that he does not need. We could use phototherapy—which might carry some risk²¹ and is costly. We could also ask his mother to discontinue breast-feeding for a day or two. The best strategy may be to do nothing.

Because the benefits of exchange transfusion seem unlikely to exceed the risks until the total bilirubin is at least 25 to 30 mg/dl, phototherapy or interruption of nursing probably should be deferred until the bilirubin approaches that range (i.e., until about 20 mg/dl).

CONCLUSIONS

In the 1950s, exchange transfusion to keep the TSB below 20 mg/dl was shown to be an effective way of preventing kernicterus in babies with erythroblastosis.^{1,18,32-34} Subsequently, several investigators showed that babies without Rh disease were less likely to develop kernicterus^{2,14,17,23,35} and could be managed more conservatively. In the late 1960s and early 1970s, however, pediatricians were troubled by the statistically significant results from the CPP^{4,15} and the more striking results of smaller studies that included many premature infants and infants with hemolysis.^{8,20,25,42} Several postmortem observations of bilirubin staining of the brain in preterm infants with low bilirubin levels^{29,55} raised new concerns about possible "brain damage" in surviving infants who had low levels of bilirubin in the newborn period.

This information was (erroneously) extrapolated to healthy full-term infants and, undoubtedly spurred on by medicolegal pressures, the exchange transfusion threshold for these infants crept downward. Then phototherapy arrived. It lowered the serum bilirubin painlessly, and pediatricians could not be blamed for assuming that it thus reduced the risk of bilirubin-induced brain damage. As we have demonstrated, however, no evidence exists to support the view that bilirubin levels of 20 mg/dl are hazardous to healthy term infants nor does any evidence suggest that phototherapy has affected cognitive or neurologic outcome in any way.⁵⁰

Past thinking about the hazards of hyperbilirubinemia in term well babies seems to have suffered from several deficiencies. First, clinicians have been misled by results that are statistically significant but clinically irrelevant. The interpretation of the data from the CPP illustrates this problem. Because of the enormous sample size, highly significant results (low *P* values) were obtained although the effect sizes were trivial. Second, many important CPP results have been published in monographs^{5,6,26} and have never been cited in the neonatal jaundice literature. Third, the implications of confounding or effect modification by prematurity and hemolysis have either not been recognized or have received insufficient attention. Premature babies and infants with hemolytic disease must be considered separately. If they are not, the relationship between bilirubin and adverse outcome in well babies is likely to be overestimated. Finally, the availability and apparent efficacy and innocuousness of phototherapy have led to its widespread use. We now see fewer babies with high bilirubin levels, and, consequently, little opportunity and less motivation exist to study the effects of such levels on cognitive or neurologic outcome.

We cannot assume, however, that 1 or 2 days of phototherapy, the interruption of nursing for 48 hours, or merely an expression of concern

about jaundice are benign events that have no adverse consequences for mothers and infants in the early postnatal period. In fact, the mere presence of neonatal jaundice (and the aura of alarm that it engenders) has negative consequences for maternal behavior and attitudes in the early, crucial months of infant development.²¹ Laboratory investigations are initiated and often repeated in normal infants with physiologic jaundice. These tests are painful for the infant, distressing to the parents, costly, and rarely enlightening.^{30,31,40}

The aggressive treatment of jaundice in term babies without hemolysis has been questioned.^{16,58} We have shown that good evidence exists that such infants are not at risk of mental or physical impairment until serum bilirubin levels rise well above 20 mg/dl. Before we treat the thousands of babies in our nurseries who daily manifest this benign form of neonatal jaundice, we should be sure that the potential benefits of our interventions outweigh the risks. Current evidence suggests that this is not the case.

ACKNOWLEDGMENT

Supported in part by grant MCJ-060573, awarded by the Bureau of Maternal and Child Health, Health Resources and Services Administration, Department of Health and Human Services, Washington, DC; and by the Robert Wood Johnson Foundation's Clinical Scholars Program.

REFERENCES

1. Allen SH, Diamond LK, Vaughn VC: Erythroblastosis fetalis: VI. Prevention of kernicterus. *Am J Dis Child*, 80:779-791, 1950
2. Bengtsson B, Verneholt J: A follow-up study of hyperbilirubinemia in healthy, full term infants without isoimmunization. *Acta Paediatr Scand* 63:70-80, 1974
3. Bjure J, Liden G, Reinard T, et al: A follow-up study of hyperbilirubinemia in full term infants without isoimmunization. *Acta Paediatr Scand* 50:437-443, 1961
4. Boggs T, Hardy J, Frazier T: Correlation of neonatal serum total bilirubin concentration and developmental status at age eight months. Preliminary report from the collaborative project. *J Pediatr* 71:553-560, 1967
5. Broman SH, Bien E, Shaughanessy P: *Low Achieving Children: The First Seven Years*. Hillsdale, NJ, Lawrence Erlbaum Associates, 1985, p 132
6. Broman SH, Nichols PL, Kennedy WA: *Preschool IQ: Prenatal and Early Developmental Correlates*. Hillsdale, NJ, Lawrence Erlbaum Associates, 1975, pp 123-142
7. Brown WR, Boon WH: Hyperbilirubinemia and kernicterus in glucose-6-phosphate dehydrogenase deficient infants in Singapore. *Pediatrics* 41:1055, 1968
8. Brown AK, Kim MH, Wu PYK, et al: Efficacy of phototherapy in prevention and management of neonatal hyperbilirubinemia. *Pediatrics* 75(suppl):393-400, 1985
9. Browner WS, Newman TB, Cummings SR, et al: Estimating sample size and power. In Hulley SB, Cummings SR (eds): *Designing Clinical Research*. Baltimore, Williams & Wilkins, 1988, pp 128-138
10. Cashore WJ: Kernicterus and bilirubin encephalopathy. *Semin Liver Dis* 8:163-167, 1988
11. Culley P, Powell J, Waterhouse J, et al: Sequelae of neonatal jaundice. *Br Med J* 3:383-386, 1970
12. Doxiadis SA, Fessas PH, Valaes T: Glucose-6-phosphate dehydrogenase deficiency: A new aetiological factor of severe neonatal jaundice. *Lancet* 1:297-301, 1961
13. Doxiadis SA, Valaes T, Karaklis A, et al: Risk of severe jaundice in glucose-6-phosphate dehydrogenase deficiency of the newborn. *Lancet* 2:1210-1213, 1964
14. Fohl E, Lombos O: The prognosis of neonatal hemolysis due to A-B-O incompatibility, without exchange transfusion. *Ann Paediatr* 203:279-287, 1964

15. Hardy JB, Peeples MO: Serum bilirubin levels in newborn infants: Distributions and associations with neurological abnormalities during the first year of life. *Johns Hopkins Med J* 128:265-272, 1971
16. Hein HA: Why do we keep using phototherapy in healthy newborns? *Pediatrics* 73:881-882, 1984
17. Holmes GE, Miller JB, Smith EE: Neonatal bilirubinemia in production of long-term neurological deficits. *Am J Dis Child* 116:37-43, 1968
18. Hsia D Y-Y, Allen FH, Gellis SS et al: Erythroblastosis fetalis: VIII. Studies of serum bilirubin in relation to kernicterus. *N Engl J Med* 247:668-671, 1952
19. Hulley SB, Cummings SR: Planning the measurement: Precision and accuracy. In Hulley SB, Cummings SR (eds): *Designing Clinical Research*. Baltimore, Williams & Wilkins, 1988, pp 37-38
20. Hyman CB, Keaster J, Hanson V, et al: CNS abnormalities after neonatal hemolytic disease or hyperbilirubinemia. *Am J Dis Child* 117:395-405, 1969
21. Kemper K, Forsyth B, McCarthy P: Jaundice, terminating breast-feeding, and the vulnerable child. *Pediatrics* 84:773-778, 1989
22. Kappas A, Drummond GS, Manola T, et al: Sn-protoporphyrin use in the management of term newborns with direct Coombs-positive ABO-incompatibility. *Pediatrics* 81:485-497, 1988
23. Killander A, Michaelsson M, Muller-Eberhard U, et al: Hyperbilirubinemia in full term newborn infants: A follow-up study. *Acta Paediatr Scand* 52:481-484, 1963
24. Killander A, Muller-Eberhard U, Sjolín, S: Indications for exchange transfusion in newborn infants with hyperbilirubinemia not due to Rh immunization. *Acta Paediatr Scand* 49:377-390, 1960
25. Johnson L, Boggs TR: Bilirubin-dependant brain damage: Incidence and indication for treatment. In Odell GB, Schaffer R, Simopoulos AP (eds): *Phototherapy in the Newborn: An Overview*. Washington, DC, National Academy of Sciences, 1974, pp 122-149
26. Lassman FM, Fisch RO, Vetter DK, et al: In La Benz PJ, La Benz ES (eds): Early correlates of speech, language and hearing. Littleton, MA, PSG Publishing, 1980
27. Lu TC, Wei H, Blackwell RQ: Increased incidence of severe hyperbilirubinemia among newborn Chinese infants with G-6-P D deficiency. *Pediatrics* 37:994-999, 1968
28. McGee DL: A program for logistic regression on the IBM PC. *Am J Epidemiol* 124:702-705, 1986
29. Maisels MJ: Neonatal jaundice. In Avery GB (ed): *Neonatology Pathophysiology and Management of the Newborn*. Philadelphia, JB Lippincott, 1987, pp 534-629
30. Maisels MJ, Gifford K: Normal serum bilirubin levels in the newborn and the effect of breast feeding. *Pediatrics* 78:837-843, 1986
31. Maisels MJ, Gifford K, Antle CE, et al: Jaundice in the healthy newborn infant: A new approach to an old problem. *Pediatrics* 81:505-511, 1988
32. Mollison PL, Cutbush M: A method of measuring the severity of a series of cases of haemolytic disease of the newborn. *Blood* 6:777-788, 1951
33. Mollison PL, Cutbush M: Haemolytic of the newborn. In Gairdner D (ed): *Recent Advances in Pediatrics*. New York, P. Blakiston & Son, 1954, p 110
34. Mollison PL, Walker W: Controlled trials of the treatment of haemolytic disease of the newborn. *Lancet* 1:429-433, 1952
35. Mores A, Fargasova I, Minarikova E: The relation of hyperbilirubinemia in newborns without isoimmunization to kernicterus. *Acta Paediatr Scand* 48:590-602, 1959
36. Naeye RL: The role of congenital bacterial infections in low serum bilirubin brain damage. *Pediatrics* 62:497-503, 1978
37. Nakamura H, Takada S, Shimabuku R, et al: Auditory and brainstem responses in newborn infants with hyperbilirubinemia. *Pediatrics* 75:703-708, 1985
38. National Institute of Child Health and Human Development: Randomized, controlled trial of phototherapy for neonatal hyperbilirubinemia. *Pediatric* 72(suppl):384-441, 1985
39. Newman TB, Browner WS, Cummings SR, et al: Designing a new study: II. Cross-sectional and case-control studies. In Hulley SB, Cummings SR (eds): *Designing Clinical Research*. Baltimore, Williams & Wilkins, 1988, pp 75-86
40. Newman TB, Easterling MJ, Goldman ES, et al: Laboratory evaluation of jaundiced newborns: Frequency, cost and yield. *Am J Dis Child*, 144:364-368, 1990

41. Newman TB, Maisels MJ: Bilirubin and brain damage: What do we do now? *Pediatrics* 83:1062-1065, 1989
42. Odell GB, Storey GNB, Rosenberg LA: Studies in kernicterus: III. The saturation of serum proteins with bilirubin during neonatal life and its relationship to brain damage at five years. *J Pediatr* 76:12-21, 1970
43. Perlman M, Fainmesser P, Sohmer H, et al: Auditory nerve-brainstem evoked responses in hyperbilirubinemic neonates. *Pediatrics* 72:658-664, 1983
44. Rosner B: *Fundamentals of Biostatistics*, ed 2. Boston, Duxbury Press, 1986, p 411
45. Ross JW, Fraser MD, Moore TD: Analytic clinical laboratory precision—state of the art for thirty-one analytes. *Am J Clin Pathol* 74:521-530, 1980
46. Rosta ZM, Bekefi D, Popper P, et al: Neonatal pathological jaundice: Seven to nine years follow up. *Acta Paediatr Acad Sci Hung* 12:317-321, 1971
47. Roth P, Polin RA: Controversial topics in kernicterus. *Clin Perinatol* 15:965-990, 1988
48. Rubin RA, Balow B, Fisch RO: Neonatal serum bilirubin levels related to cognitive development at ages 4 through 7 years. *J Pediatr* 94:601-604, 1979
49. Sansone G, Perroni L, Yoshida A: Glucose-6-phosphate dehydrogenase variants from Italian subjects associated with severe neonatal jaundice. *Br J Haematol* 31:159-165, 1975
50. Scheidt PC, Bryla DA, Nelson KB, et al: Phototherapy for neonatal hyperbilirubinemia: Six year follow-up of the NICHD clinical trial. *Pediatrics*, 1990, in press
51. Scheidt PC, Mellits ED, Hardy JB, et al: Toxicity to bilirubin in neonates. Infant development during first year in relation to maximum neonatal serum bilirubin concentration. *J Pediatr* 91:292-297, 1977
52. Schreiner RL, Glick MR: Interlaboratory bilirubin variability. *Pediatrics* 69:277-281, 1982
53. Tan KL: Phototherapy for neonatal jaundice in erythrocyte glucose-6-phosphate dehydrogenase deficient infants. *Pediatrics* 59:1023-1026, 1977
54. Trolle D: Discussion of the advisability of performing exchange transfusion in neonatal jaundice of unknown aetiology. *Acta Paediatr (Stockholm)* 50:392, 1961
55. Valaes T: Estimation and clinical significance of bilirubin binding status. In Rubaltelli FF, Jori G (eds): *Neonatal jaundice: New trends in phototherapy*, New York, Plenum Press, 1984, pp 79-93
56. Valaes T, Kipourou K, Petmezaki S, et al: Effectiveness and safety of prenatal phenobarbital for the prevention of neonatal jaundice. *Pediatr Res* 14:947-952, 1980
57. Vohr BR, Lester B, Rapisardi G, et al: Abnormal brain-stem function (brain-stem auditory evoked response) correlates with acoustic cry features in term infants with hyperbilirubinemia. *J Pediatr* 115:303-308, 1989
58. Watchko JF, Oski FA: Bilirubin 20 mg/dl = vigintiphobia. *Pediatrics* 71:660-663, 1983
59. Watkinson LR, St John A, Pemberthy LA: Investigation into paediatric bilirubin analyses in Australia and New Zealand. *J Clin Pathol* 35:52-58, 1982

Address reprint requests to

Thomas B. Newman, MD, MPH
 Division of General Pediatrics
 P.O. Box 0314
 University of California
 San Francisco, CA 94143