Performance Metrics After Changes in Screening Protocol for Congenital Hypothyroidism

WHAT'S KNOWN ON THIS SUBJECT: Significant variation in congenital hypothyroidism screening operations/performance has been observed in the United States. The origin of this variation remains unknown, in part because of a lack of evaluation. Accordingly, debates persist about optimal screening operations including laboratory testing methods.

WHAT THIS STUDY ADDS: Four distinct screening protocols applied to Michigan resident infants are compared in detecting congenital hypothyroidism overall and specific to cases characterized by high initial thyrotropin concentrations thought to have a more severe form of the disease.

abstract

OBJECTIVE: To evaluate Michigan newborn screening for congenital hypothyroidism (CH) protocol changes.

METHODS: This population-based study includes infants born and screened in Michigan (January 1, 1994–June 30, 2010). Screening performance is compared across 4 periods defined by the dried blood spot testing method: (1) thyroxine (T4) with backup thyrotropin, (2) tandem T4 and thyrotropin, (3) primary thyrotropin testing without serial testing, and (4) primary thyrotropin plus serial testing for births weighing <1800 g. Logistic regression is used to test for differences across periods.

RESULTS: Thyrotropin testing exhibited greater specificity overall and greater likelihood of detection with serial testing relative to primary T4 testing. Tandem T4 and thyrotropin testing appeared more sensitive relative to other protocols, yet it produced significantly more false-positives, and detection may have been affected by overdiagnosis and misclassification. Central CH was no longer detected once T4 testing ceased.

CONCLUSIONS: Primary thyrotropin plus serial testing for infants at risk for later rising thyrotropin outperformed other newborn screening strategies for classic CH, although 2 false-negatives occurred among normal birth weight infants admitted to the NICU during this testing period. Tandem T4 and thyrotropin screening outperformed other strategies for detection of both classic and central CH combined, although it is associated with increased operating costs. Additional research is necessary to weigh the benefits of increased sensitivity against additional program operating costs. *Pediatrics* 2012;130:e1252–e1260

AUTHORS: Steven J. Korzeniewski, PhD,^{a,b} Violanda Grigorescu, MD, MSPH,^c Mary Kleyn, MS,^c William Young, PhD,^c Gretchen L. Birbeck, MD,^d David Todem, PhD,^d Roberto Romero, MD, DMedSci,^a Tinnakorn Chaiworapongsa, MD,^{a,b} and Nigel Paneth, MD, MPH^d

^aPerinatology Research Branch, National Institute of Child Health and Human Development, National Institutes of Health, Department of Health and Human Services, Detroit, Michigan;

^bDepartment of Obstetrics and Gynecology, Wayne State University, Detroit, Michigan;

^cMichigan Department of Community Health, Lansing, Michigan; and

^dDepartment of Epidemiology, Michigan State University, East Lansing, Michigan

KEY WORDS

newborn screening, performance evaluation, congenital hypothyroidism

ABBREVIATIONS

CH—congenital hypothyroidism

DOB—date of birth

FPR—false-positive rate

NBS—newborn screening

PPV—positive predictive value

T4—thyroxine

Each author made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; participated in drafting the article or revising it critically for important intellectual content; and provided final approval of the version to be published.

www.pediatrics.org/cgi/doi/10.1542/peds.2011-3340

doi:10.1542/peds.2011-3340

Accepted for publication Jun 20, 2012

Address correspondence to Steven J. Korzeniewski, PhD, Department of Obstetrics and Gynecology, Wayne State University School of Medicine, Hutzel Women's Hospital, 4 Brush—Office 4817, 3990 John R St, Detroit, MI 48201. E-mail: sKorzeni@med.wayne.edu

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

Copyright © 2012 by the American Academy of Pediatrics

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

FUNDING: This research was supported in part by the Perinatology Research Branch, Division of Intramural Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Department of Health and Human Services. Funded by the National Institutes of Health (NIH).

Newborn screening (NBS) for congenital hypothyroidism (CH), a clinically defined group of thyroid disorders observed at birth, began in the mid-1970s after the development of a radioimmunoassay capable of measuring thyroxine (T4) in dried blood spotted on filter paper.^{1–5} Based on findings from the first million infants screened, the NBS Committee of the American Thyroid Association recommended broad establishment and expansion of NBS programs for CH in 1977.⁶ By 1992, it was estimated that 50 million infants were screened annually for CH worldwide.⁷ NBS programs around the world initially reported detection rates ranging from 1:3000 to 1:4000 infants screened and a typical 2:1 ratio of female to male cases.^{8–11} More recently, US NBS programs have reported an increase in the birth prevalence of CH from 1:3985 in 1987 to 1:2274 in 2002 not fully explained by changes in laboratory methods or potential misclassification of transient disease; significant interstate variation has also been observed.^{12–16} The origin of this variation remains largely unknown, perhaps because there has been a lack of emphasis on evaluating screening system components.¹⁷ Accordingly, debates persist about optimal CH screening operations, particularly dried blood spot testing methods.

Previous evidence of the comparative effectiveness of dried blood spot testing protocols for CH NBS is heterogeneous. Greater sensitivity and specificity have been reported among primary thyrotropin relative to T4 testing programs and vice versa.^{15,18–24} Several studies estimated that 4% to 10% of cases missed by primary T4 testing are appropriately detected by primary thyrotropin testing^{20–23}; others conclude that primary thyrotropin testing fails to detect cases of central CH and those exhibiting a later rise in thyrotropin.^{18,19} Serial thyrotropin testing protocols that

rescreen selected infants generally during the first month of life have emerged to address later rising thyrotropin^{25–27}; however, detection of central CH remains an issue. Estimates of the birth prevalence of central CH range widely (~1:20 000–1:125 000 live births), and some believe it is adequately detected clinically amid diagnosis of concomitant pituitary hormone deficiencies. Others note that although it is true that most (~75%) cases of central hypothyroidism also have pituitary hormone deficiencies,²⁸ diagnosis of either condition can often be delayed beyond 3 months of age and may result in severe hypoglycemia, neonatal hepatitis, or death.^{29–31}

Heterogeneity of previous evidence is difficult to interpret in the context of interprogram variation in screening protocols, performance, and population characteristics.³² In an attempt to investigate the impact of changes in NBS for CH among a fixed population, we assessed NBS for CH performance metrics in Michigan during 4 successive periods in which different dried blood spot testing protocols were used. This study adds to previous literature by (1) comparing the effectiveness of 4 distinct screening protocols in a reasonably stable and homogenous population of infants, (2) reporting findings generated in a program that collects virtually all initial bloodspot specimens between 24 and 36 hours of life, and (3) comparing NBS protocols based on their ability to detect cases characterized by high initial thyrotropin concentrations (100 uIU/mL) who are thought to have a more severe form of CH in addition to overall CH.

METHODS

Study Design and Participants

This population-based retrospective cohort study was approved by the Michigan Department of Community

Health Institutional Review Board and includes Michigan resident infants born and screened in Michigan from January 1, 1994, through June 30, 2010. The primary exposure of interest is the method of dried blood spot testing defined based on infant date of birth (DOB) as follows: (1) T4 backup thyrotropin testing (DOB: 1/1/1994–12/31/1997); (2) tandem T4 and thyrotropin testing for all infants, no serial testing (DOB: 1/1/1998–9/30/2003); (3) primary thyrotropin testing, no serial testing (DOB: 10/1/2003–2/28/2007); and (4) primary thyrotropin testing, serial testing for infants weighing <1800 g at birth (DOB: 3/1/2007–6/30/2010). T4 backup thyrotropin testing involves making referrals for confirmatory testing based on thyrotropin determinations obtained from dried blood spots only in newborns whose T4 concentrations are below the 10th centile. Tandem T4 and thyrotropin testing involves making referrals for confirmatory testing based on either low T4 or elevated thyrotropin concentration measured in newborn dried blood. Primary thyrotropin testing involves making referrals for confirmatory testing based only on thyrotropin concentration; the addition of serial testing involves rescreening among infants at elevated risk of later rising thyrotropin.

Confirmatory testing is usually based on serum tests of venipuncture blood samples combined with some measure of binding proteins (ie, T3 resin uptake) used to differentiate free (active) from total T4.^{24,64} Blood samples for confirmatory testing are ideally obtained ~2 to 3 weeks of life when the upper range of thyrotropin falls to ~10 mU/L. Reference ranges for free T4, total T4, and thyrotropin concentrations measured in serum at 2 to 4 weeks of life are ~10 to 26 pmol/L, 90 to 206 nmol/L, and <10 mU/L, respectively.²² Infants having ≥2 serum thyrotropin concentrations >20

mU/L are expected to have permanent primary CH.¹³ If a defect in thyroid hormone synthesis is suspected, perchlorate washout testing is sometimes performed to test the ability of the thyroid to transform iodine into organically bound iodine.⁶⁵ Other tests including scintigraphy and ultrasound are also useful during the process of diagnosing CH. Table 1 reports cutoff values used in referring infants for confirmatory testing over time. Outcomes of interest include screening performance metrics: detection rate, false-positive rate (FPR), positive predictive value (PPV), sensitivity, and specificity.

Data

Demographic and perinatal information collected on the NBS, laboratory screening results, and medical management data were used to identify and characterize infants screened from January 1, 1994 through June 30, 2010. Aside from rescreens because of early

specimen collection, infants identified by newborn dried blood spot screen for additional testing are considered screen positive; those who are classified as CH and are treated at the conclusion of confirmatory testing are considered diagnosed cases. Reports actively and passively ascertained from pediatric endocrinologists by the NBS Follow-up Program are used in this study to identify false-negative screening results.

Analysis

Descriptive and analytical techniques include tabulation and trending of newborn characteristics by NBS outcomes of interest during 4 exposure periods. Logistic regression analysis is used to investigate whether the overall likelihood of detection, likelihood of severe CH detection, and likelihood of false-positive determination changed significantly across periods after adjusting for differences in the distribution of selected newborn demographic and

perinatal characteristics. Cases are categorized as severe CH if their initial thyrotropin concentration reached or exceeded 100 uIU/mL based on the work of Mitchell et al.³³ Adjusted models include covariates that are both significantly associated with the dependent variable (overall detection or severe case detection) and varied significantly during the 4 exposure periods. We were unable to assess area under the receiver operating characteristic curve associated with each protocol because of the lack of analyte concentration data among normal screens during T4 testing periods.

RESULTS

More than 2 million infants are included; Table 2 reports the distributions of demographic and perinatal characteristics across the 4 exposure periods. Population characteristics did not meaningfully differ over time, although, due to the large sample size, observed differences were statistically significant.

Table 3 reports screening performance metrics by dried blood spot testing protocol. During the T4 backup thyrotropin testing period, the detection rate, positive predictive value, and specificity were each less than observed during primary thyrotropin testing periods, both with and without serial testing. Alternatively, the FPR was more than twofold greater during the primary T4 relative to primary thyrotropin testing periods. The greatest rate of overall detection was observed during the tandem T4 and thyrotropin testing period (1:1271); however, the FPR (4.45%) was far greater than during other periods of observation. Accordingly, the PPV and specificity were significantly less during the tandem T4 and thyrotropin testing period compared with others observed in this study. Of note, the expected gender dimorphism of more female than male cases was reversed

TABLE 1 Dried Blood Spot Testing Protocols, Cutoff Values, and Associated Determinations Applied by Michigan NBS for Congenital Hypothyroidism, 1977–2010

Time Period	Test Method	T4 (ng/dL)	Age at Specimen Collection	Thyrotropin (uIU/mL)	Determination ^a
1977–1997	T4 follow-up thyrotropin	<10th centile	All ages	>20	Positive
1998–2003	Tandem T4 and thyrotropin	<5.0	All ages	23–49	Borderline positive
2004–2010	Primary thyrotropin	NA	<24 h	>49	Strong positive
				≥50	Strong positive
			24–36 h	<50	Early specimen
				<33	Negative
				33–49	Borderline positive
				≥50	Strong positive
			37 h–6 d	<25	Negative
				25–49	Borderline positive
				≥50	Strong positive
			7–31 d	<13	Negative
13–49	Borderline positive				
≥50	Strong positive				
>31 d	≤10	Negative			
	>10	Positive			

NA, not applicable.
^a Positive or strong positive determination results in referral for confirmatory testing. Early specimen and borderline positive determinations result in repeat dried blood spot screen; after an initial borderline-positive determination, if the repeat dried blood spot screen is also positive, then the newborn is referred for confirmatory testing. Daily percentiles were used to refer children for confirmatory testing from 1977 to 1997. Serial testing was implemented for infants in the NICU born weighing <1800 g March 1, 2007.

TABLE 2 Newborns Screened by Selected Demographic and Perinatal Characteristics and by Dried Blood Spot Testing Method, Michigan NBS, January 1, 1994 through June 30, 2010

Population Segment	T4 Backup Thyrotropin		Tandem T4 and Thyrotropin		Thyrotropin, No Serial Testing		Thyrotropin, Plus Serial Testing	
	(1/1994–12/1997)		(1/1998-9/2003)		(10/2003–2/2007)		(3/2007–6/2010)	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Race								
White	386459	73.37	511212	72.23	281854	71.90	253137	70.59
Black	103361	19.62	129240	18.26	73688	18.80	70611	19.69
Other	36929	7.01	67327	9.51	36474	9.30	34845	9.72
Gender								
Female	260569	48.79	364919	48.75	211349	48.85	193165	48.85
Male	273445	51.21	383625	51.25	221266	51.15	202298	51.15
Multiple Birth								
No	526879	97.04	729276	96.63	417431	96.49	381470	96.46
Yes	16066	2.96	25446	3.37	15184	3.51	13993	3.54
Birth Weight								
<750 g	2249	0.43	4410	0.58	1199	0.28	144	0.29
750–1499 g	643.7	1.23	8293	1.10	4690	1.10	4352	1.12
1500–2499 g	35301	6.73	47864	6.35	28308	6.65	25778	6.62
>2500 g	480752	91.62	693477	91.97	391351	91.96	358355	91.97
Length of gestation								
<28 wk	—	—	—	—	2327	0.57	2184	0.57
28–37 wk	—	—	—	—	40212	9.85	36115	9.42
>37 wk	—	—	—	—	365499	89.57	344903	90.01
Small for gestational age								
No	—	—	—	—	318936	90.44	295320	90.23
Yes	—	—	—	—	33698	9.56	31964	9.77
NICU								
No	—	—	—	—	387543	89.58	353966	89.51
Yes	—	—	—	—	45072	10.42	41497	10.49
Total	542945	100	754722	100	432615	100	395463	100

Missing data are as follows: race, $n = 140\ 620$; gender, $n = 22\ 140$; birth weight, $n = 31\ 786$; gestational age, $n = 36\ 843$. —, NICU admission and gestational age data were not recorded on the NBS card before October 1, 2003. Percentages reported are column based.

only during the tandem T4 and thyrotropin testing period, suggesting potential misclassification; otherwise, more female than male infants were diagnosed as expected.

Primary thyrotropin testing protocols were more specific than primary T4 testing protocols, yielding greater PPVs; however, the detection rate observed during primary thyrotropin testing periods is less than was observed during the tandem T4 and thyrotropin

testing period. Overall, primary thyrotropin testing with serial testing for infants born weighing <1800 g yielded fewer false-positive results, a greater PPV, and greater overall detection than either primary thyrotropin testing without serial testing or primary T4 backup thyrotropin testing protocols. However, the 2 false-negative results observed during this study occurred during the primary thyrotropin plus serial testing period among infants

admitted to the NICU who had later rising thyrotropin but were not included in the serial testing protocol due to their normal birth weights. A single case of central hypothyroidism was detected during the T4 backup thyrotropin testing period (1:542 945), 6 of such cases were detected during the tandem T4 and thyrotropin testing period (1:125 787), and none were detected during either primary thyrotropin testing periods.

TABLE 3 NBS Results and Performance Metrics, Michigan, January 1, 1994–June 30, 2010

Protocol	Screened, <i>N</i>	Screened Positive, <i>n</i>	Diagnosed, <i>n</i>	False-Negative, <i>n</i>	Detection Rate	PPV, %	FPR, %	Se, %	Sp, %
T4 backup thyrotropin	542 945	10 454	241	0	1:2253	2.31	1.88	100	98.12
Tandem T4 and thyrotropin	754 722	34 172	594	0	1:1271	1.74	4.45	100	95.55
Thyrotropin, no serial testing	432 615	3601	225	0	1:1923	6.25	0.78	100	99.22
Thyrotropin, plus serial testing	395 463	2375	259	2	1:1527	10.90	0.54	99.2	99.46

Detection rate is the number of live births screened to identify a single case of congenital hypothyroidism. PPV, positive predictive value; FPR, false-positive rate; Se, sensitivity; Sp, specificity.

Although the overall rate of CH detection and subsequent screening performance metrics varied considerably by screening protocol period, the birth prevalence of severe CH, characterized by having an initial thyrotropin concentration >100 uIU/mL, was far more stable (Fig 1). The number of severe CH cases detected per 100 000 live births screened increased after the introduction of thyrotropin into the dried blood spot testing protocol relative to the T4 backup thyrotropin testing period and remained relatively stable across the tandem T4 and thyrotropin and primary thyrotropin testing periods (with and without serial testing).

Overall, after adjusting for potential confounding factors (race, gender, twin status, birth weight), tandem T4 and thyrotropin testing and primary thyrotropin plus serial testing for infants born weighing <1800 g are associated with a 89% and 58% increase in the odds of detection respectively compared with primary T4 backup thyrotropin testing. Primary thyrotropin testing was more specific than primary T4 testing, yet was associated with a greater likelihood of detection only after introduction of serial testing.

Although tandem T4 and thyrotropin testing was associated with a near twofold increase in overall detection compared with primary T4 testing, it

was also associated with a near threefold increase in the rate of false-positives, as shown in Table 4. Alternatively, the FPR was significantly reduced during both primary thyrotropin testing periods relative to the primary T4 backup thyrotropin and tandem testing periods in both crude (unadjusted) and adjusted models. To compare the trade-offs of tandem thyrotropin and T4 testing verse primary thyrotropin plus serial testing for infants born weighing <1800 g, we applied the detection rates and FPRs to a hypothetical birth population of 125 000 infants and estimated that an additional 297 false-positive determinations would be incurred for each additional case detected if Michigan were to switch from primary thyrotropin plus serial testing back to tandem T4 and thyrotropin testing for all births.

As indicated in Table 5, the crude likelihood of severe CH detection was greatest during the thyrotropin plus serial testing period and was significantly elevated in each screening protocol period relative to the T4 backup thyrotropin testing strategy. After adjustment for race and gender distributions, the difference in likelihood of severe CH detection between T4 backup thyrotropin and primary thyrotropin without serial testing protocols was not statistically significant.

Severe CH cases were 38% and 35% more likely to be detected during tandem T4 and thyrotropin and primary thyrotropin plus serial testing periods respectively relative to the T4 backup thyrotropin testing period after adjusting for race and gender distributions.

DISCUSSION

Although the overall detection rate was greatest during the tandem T4 and thyrotropin testing period in this study, this finding is likely affected by misclassification and overdiagnosis based on the elevated birth prevalence, reversal of the expected gender dimorphism, and stable rate of severe CH observed in this period relative to primary thyrotropin testing periods. Furthermore, a surprising 72% of cases detected by primary thyrotropin exhibited normal T4 concentrations, far greater than the expected 4% to 10%,^{20–23} suggesting that cases of hyperthyrotropinemia may have been classified and treated as CH during this period. Primary thyrotropin testing plus serial testing among infants born <1800 g yielded fewer false-positives and accordingly had lesser operating costs than either tandem T4 and thyrotropin or primary T4 backup thyrotropin protocols. Primary thyrotropin plus serial testing was also associated with a greater likelihood of detection relative to primary T4 testing and was equally able to detect severe CH relative to the tandem testing approach, although no cases of central CH were detected during this period.

On average, 1 case of central CH was detected per year in Michigan before removing T4 from the NBS protocol; none were detected after. It is possible that ≥ 1 cases of central CH was missed by primary thyrotropin testing strategies and perhaps not identified because of mortality or migration before clinical detection or not reported

FIGURE 1

Birth prevalence of overall and severe (initial thyrotropin >100 uIU/mL) congenital hypothyroidism by newborn screening method, Michigan January 1, 1994 through June 30, 2010. Dark gray indicates overall CH (all cases). Light gray indicates severe CH.

TABLE 4 Magnitude of Association Between Congenital Hypothyroidism Detection, False-Positive Screening Determination, and Dried Blood Spot Testing Protocol, Michigan NBS, January 1, 1994–June 30, 2010

Population Segment	CH Detection				False-Positive Screen			
	Crude		Adjusted		Crude		Adjusted	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Protocol								
T4 backup thyrotropin ^a	1		1		1		1	
Tandem T4 and thyrotropin	1.77	1.53–2.06	1.89	1.61–2.22	2.43	2.38–2.48	2.85	2.78–2.92
Thyrotropin, no serial testing	1.17	0.98–1.41	1.19	0.98–1.46	0.41	0.40–0.43	0.48	0.46–0.50
Thyrotropin, plus serial testing	1.48	1.24–1.76	1.58	1.30–1.91	0.28	0.27–0.29	0.33	0.32–0.35
Race								
White ^a	1	—	1	—	1	—	1	—
Black	0.97	0.83–1.12	0.86	0.73–1.00	1.54	1.50–1.57	1.37	1.34–1.40
Other	1.62	1.37–1.91	1.53	1.28–1.82	0.91	0.88–0.94	0.87	0.84–0.91
Gender								
Male ^a	1	—	1	—	1	—	1	—
Female	1.12	1.00–1.25	1.09	0.97–1.22	0.78	0.76–0.79	0.76	0.74–0.77
Twin								
No ^a	1	—	1	—	1	—	1	—
Yes	2.08	1.68–2.58	1.23	0.96–1.57	2.10	2.03–2.18	1.09	1.04–1.14
Birth Wt								
<750 g	9.52	7.04–12.9	5.81	6.33–12.26	12.04	11.4–12.7	10.5	9.87–11.1
750–1499 g	5.48	4.28–7.02	5.81	4.46–7.56	8.38	8.06–8.70	8.48	8.13–8.85
1500–2499 g	2.08	1.75–2.47	2.08	1.72–2.52	2.24	2.17–2.30	2.17	2.10–2.24
≥2500 g ^a	—	—	—	—	—	—	—	—

Table reports the likelihood of CH detection and separately false-positive screening determination among population segments compared with referents expressed as an odds ratio (OR). Odds ratio confidence intervals (CIs) that do not encompass 1 are indicative of a significant difference compared with the referent category. Adjusted models include all covariates listed in the table. Missing data are as follows: race, $n = 140\ 620$; gender, $n = 22\ 140$; birth wt, $n = 31\ 786$.

^a Referents.

because of our reliance on passive surveillance of false-negatives. It remains unclear how the additional operating costs associated with tandem thyrotropin and T4 testing compare with the benefit of early central CH detection, although additional research is necessary to quantify this benefit.

Additional investigation is also necessary to determine whether there is benefit to increased detection of marginal cases including hyperthyrotropinemia/subclinical CH, and hypothyroxinemia, particularly in the context of significant increases in the number of diagnoses in the United States over past 20 years. Currently, little evidence exists about the cognitive outcomes of permanent or transient forms of hyperthyrotropinemia and subclinical hypothyroidism.^{34–37} Two small studies reported an average decrement of 7 to 8 IQ points among children having hyperthyrotropinemia compared with euthyroid children;^{38,39} another

reported subclinical hypothyroidism after age 5 years among such cases.⁴⁰ Alternatively, other small studies have reported normal mental and physical development among untreated hyperthyrotropinemia and subclinical CH cases.^{41–44} Several investigators have also reported potential harm including iatrogenic hyperthyroidism associated with treatment of hyperthyrotropinemia patients.^{45,46}

It is similarly unclear whether cases of hypothyroxinemia, a condition common among preterm infants and characterized by low T4 concentrations and normal thyrotropin concentrations not associated with CH, should be treated. Although NBS programs have traditionally considered positive screening results associated with hypothyroxinemia as being false, evidence suggests that these children are at elevated risk for neurodevelopmental disorders⁴⁷ and developmental delay⁴⁸; trials are underway to determine

whether there is a benefit from treatment, results may have implications for future NBS operations.^{49,50}

Future research efforts would be greatly advanced by application of a standardized operational case definition for CH across NBS programs, similar to efforts made in surveillance for cerebral palsy in Europe.⁵¹ Absent a standardized operational case definition, it is difficult to make meaningful comparisons between and within screening programs over time. This definition should lay out the criteria for diagnosing CH in terms of necessary tests and how to interpret them and should attempt to differentiate classic CH from other congenital thyroid abnormalities using operational terms. Standardized age-adjusted analyte thresholds are recommended for both dried blood and serum measurements. It is also recommended that all suspected cases of CH undergo thyroid imaging to facilitate differentiation of likely transient from permanent

TABLE 5 Magnitude of Association Between Severe Congenital Hypothyroidism Detection and Dried Blood Spot Testing Protocol, Michigan NBS, January 1, 1994–June 30, 2010

Population Segment	Detection of Severe CH ^a			
	Crude		Adjusted	
	OR	95% CI	OR	95% CI
Serial Testing				
T4 backup thyrotropin ^b	1		1	
Tandem T4 and thyrotropin	1.35	1.07–1.70	1.38	1.09–1.75
Thyrotropin, no serial testing	1.35	1.05–1.75	1.25	0.95–1.65
Thyrotropin, plus serial testing	1.51	1.17–1.94	1.35	1.02–1.77
Race				
White ^b	1	—	1	—
Black	0.34	0.24–0.47	0.35	0.25–0.49
Other	1.35	1.04–1.74	1.34	1.03–1.74
Sex				
Male ^b	1	—	1	—
Female	1.91	1.60–2.27	1.88	1.56–2.25
Multiple Birth				
No ^b	1	—	—	—
Yes	1.17	0.76–1.78	—	—
Birth Weight				
<750 g	1.73	0.65–4.64	—	—
750–1499 g	1.15	0.55–2.42	—	—
1500–2499 g	1.31	0.97–1.77	—	—
≥2500 g ^b	—	—	—	—

Table reports the likelihood of CH detection and separately false-positive screening determination among population segments compared with referents expressed as an odds ratio (OR). Odds ratio confidence intervals (CIs) that do not encompass 1 are indicative of a significant difference compared with the referent category. The adjusted model includes only race and gender as covariates because multiple birth and birth weight were not significantly associated with detection of severe CH and their inclusion did not alter the point estimate (OR) of the primary exposure variable by >10%. Missing data are as follows: race, *n* = 140 620; gender, *n* = 22 140; birth wt, *n* = 31 786.

^a Initial thyrotropin concentration ≥100 uIU/mL.

^b Referents.

cases. Finally, expansion of long term follow-up and data collection activities including neurodevelopmental assessment would also facilitate future investigation of cost-benefit.

This study is limited by missing data, although it appeared to occur at random based on similar distributions among tabulations by overall CH detection, severe CH detection, and false-positive screening determination. The small proportion of cases that underwent thyroid imaging (15%) hindered our ability to investigate further whether transient or milder forms of CH were more likely to be detected during any

of the observed exposure periods. Our definition of severe CH is also imprecise; 56% of infants included in this study who exhibited an initial thyrotropin concentration ≥100 uIU/mL were not diagnosed as CH. However, use of Mitchell et al.'s definition of severe CH revealed an interesting trend in detection across protocols and led us to similarly believe it is unlikely that the true birth prevalence of classic CH has increased over time. Our findings are also negatively affected by reliance on passive reporting to identify false-negative screening results; accordingly, our results per false-negative

determinations should be interpreted as a minimum. Finally, this study is limited by the lack of universal long-term follow-up beyond age 3 years, meaning we are unable to differentiate permanent from transient CH.

CONCLUSIONS

Overall, our findings suggest that primary thyrotropin plus serial testing for infants at risk for later rising thyrotropin is an effective NBS strategy for classic CH (characterized by elevated thyrotropin and low T4), although the 2 false-negatives observed in this study occurred among normal birth weight infants admitted to the NICU during this period due to later rising thyrotropin. Michigan now rescreens all children admitted to the NICU at 30 days of life or discharge in lieu of retesting at 14 days and again at 28 days of life only among children born weighing <1800 g. Additional evaluations are underway to determine if the revised serial testing protocol adequately addressed the false-negatives observed in this study. Tandem T4 and thyrotropin screening outperformed other strategies for detection of both classic and central CH combined, although it is associated with increased operating costs per additional laboratory infrastructure and increased false-positive determinations primarily among preterm infants. Additional research is necessary to weigh the benefits of increased sensitivity against additional program operating costs; this research should support future guidelines directly addressing whether central CH should be included in the recommended panel of NBS conditions.

REFERENCES

1. Alm J, Larsson A, Zetterström R. Congenital hypothyroidism in Sweden. Incidence and age at diagnosis. *Acta Paediatr Scand*. 1978;67(1):1–3
2. Klein AH, Agustin AV, Foley TP Jr. Successful laboratory screening for congenital hypothyroidism. *Lancet*. 1974;2(7872):77–79
3. Dussault JH, Coulombe P, Laberge C, Letarte J, Guyda H, Khoury K. Preliminary report on a mass screening program for neonatal hypothyroidism. *J Pediatr*. 1975;86(5):670–674

4. Klein AG, Foley TP Jr. Letter: Screening for hypothyroidism. *J Pediatr*. 1975;87(4):668–8
5. LaFranchi SH, Murphey WH, Foley TP Jr, Larsen PR, Buist NR. Neonatal hypothyroidism detected by the Northwest Regional Screening Program. *Pediatrics*. 1979;63(2):180–191
6. Fisher DA, Burrow GN, Dussault JH, et al. Recommendations for screening programs for congenital hypothyroidism. Report of a committee of the American Thyroid Association. *Am J Med*. 1976;61(6):932–934
7. Delange F. Neonatal screening for congenital hypothyroidism: results and perspectives. *Horm Res*. 1997;48(2):51–61
8. Fisher DA. Second International Conference on Neonatal Thyroid Screening: progress report. *J Pediatr*. 1983;102(5):653–654
9. Lafranchi S. Congenital hypothyroidism—a newborn screening success story? *Endocrinologist*. 1994;4(6):477–486
10. Klett M. Epidemiology of congenital hypothyroidism. *Exp Clin Endocrinol Diabetes*. 1997;105(suppl 4):19–23
11. LaFranchi S. Congenital hypothyroidism: etiologies, diagnosis, and management. *Thyroid*. 1999;9(7):735–740
12. Shapira SK, Lloyd-Puryear MA, Boyle C. Future research directions to identify causes of the increasing incidence rate of congenital hypothyroidism in the United States. *Pediatrics*. 2010;125(suppl 2):S64–S68
13. Parks JS, Lin M, Grosse SD, et al. The impact of transient hypothyroidism on the increasing rate of congenital hypothyroidism in the United States. *Pediatrics*. 2010;125(suppl 2):S54–S63
14. Olney RS, Grosse SD, Vogt RF Jr. Prevalence of congenital hypothyroidism—current trends and future directions: workshop summary. *Pediatrics*. 2010;125(suppl 2):S31–S36
15. Hertzberg V, Mei J, Therrell BL. Effect of laboratory practices on the incidence rate of congenital hypothyroidism. *Pediatrics*. 2010;125(suppl 2):S48–S53
16. Hinton CF, Harris KB, Borgfeld L, et al. Trends in incidence rates of congenital hypothyroidism related to select demographic factors: data from the United States, California, Massachusetts, New York, and Texas. *Pediatrics*. 2010;125(suppl 2):S37–S47
17. Therrell BL Jr, Schwartz M, Southard C, Williams D, Hannon WH, Mann MY, PEAS Organizing and Working Groups. Newborn Screening System Performance Evaluation Assessment Scheme (PEAS). *Semin Perinatol*. 2010;34(2):105–120
18. Mandel SJ, Hermos RJ, Larson CA, Prigoghin AB, Rojas DA, Mitchell ML. Atypical hypothyroidism and the very low birth-weight infant. *Thyroid*. 2000;10(8):693–695
19. Zamboni G, Zaffanello M, Rigon F, Radetti G, Gaudino R, Tatò L. Diagnostic effectiveness of simultaneous thyroxine and thyroid-stimulating hormone screening measurements. Thirteen years' experience in the Northeast Italian Screening Programme. *J Med Screen*. 2004;11(1):8–10
20. Wang ST, Pizzolato S, Demshar HP. Diagnostic effectiveness of TSH screening and of T4 with secondary TSH screening for newborn congenital hypothyroidism. *Clin Chim Acta*. 1998;274(2):151–158
21. Tuerck JM, Miyahira R, Skeels M, Rien L, Sesser D, Buist NR. Newborn screening strategies: routine second tests—Oregon. In: Pass KA, Levy HL, eds. *Early Hospital Discharge: Impact on Newborn Screening*. Atlanta, GA: CORN; 1995:201–212
22. Therrell B. Second testing in newborn screening program in the US. In: Pass KA, Levy HL, eds. *Early Hospital Discharge: Impact on Newborn Screening*. Atlanta, GA: CORN; 1995:75–86
23. Baumgartner JH. Screening for primary hypothyroidism in Missouri, using a thyroxine and TSH assay for all specimens: a five year experience. In: Therrell B, Aldis BG, editors. *Proceedings of the 11th National Neonatal Screening Symposium*; September 12–16, 1995; Corpus Christi, TX. pp. 39–42
24. Madison LD, LaFranchi S. Screening for congenital hypothyroidism: current controversies. *Curr Opin Endocrinol Diabetes Obes*. 2005;12(1):36–41
25. Larson C, Hermos R, Delaney A, Daley D, Mitchell M. Risk factors associated with delayed thyrotropin elevations in congenital hypothyroidism. *J Pediatr*. 2003;143(5):587–591
26. Silva SAB, Chagas AJ, Goulart EMA, et al. Screening for congenital hypothyroidism in extreme premature and/or very low birth weight newborns: the importance of a specific protocol. *J Pediatr Endocrinol Metab*. 2010;23(1-2):45–52
27. Tylek-Lemańska D, Kumorowicz-Kopiec M, Starzyk J. Screening for congenital hypothyroidism: the value of retesting after four weeks in neonates with low and very low birth weight. *J Med Screen*. 2005;12(4):166–169
28. van Tijn DA, de Vijlder JJM, Verbeeten B Jr, Verkerk PH, Vulsma T. Neonatal detection of congenital hypothyroidism of central origin. *J Clin Endocrinol Metab*. 2005;90(6):3350–3359
29. Doeker B, Andler W. Congenital hypothyroidism: causes for delayed initiation of treatment [in German]. *Klin Padiatr*. 1999;211(3):161–164
30. Vlusma T, Delemarre HA, de Muinck Keizer SMPF, et al. Detection and classification of congenital thyrotropin deficiency in the Netherlands. In: *The Thyroid Gland, Environment and Autoimmunity* (International Congress Series). Amsterdam: Excerpta Medica; 1989:343–346
31. Hintz RL. Eternal vigilance—mortality in children with growth hormone deficiency. *J Clin Endocrinol Metab*. 1996;81(5):1691–1692
32. Pass KA, Neto EC. Update: newborn screening for endocrinopathies. *Endocrinol Metab Clin North Am*. 2009;38(4):827–837
33. Mitchell ML, Hsu H-W, Sahai I, and the Massachusetts Pediatric Endocrine Work G. The increased incidence of congenital hypothyroidism: fact or fancy? *Clin Endocrinol (Oxf)*. 2011;75(6):806–810
34. Rose SR, Brown RS, Foley T, et al; American Academy of Pediatrics; Section on Endocrinology and Committee on Genetics, American Thyroid Association; Public Health Committee, Lawson Wilkins Pediatric Endocrine Society. Update of newborn screening and therapy for congenital hypothyroidism. *Pediatrics*. 2006;117(6):2290–2303
35. Krude H, Blankenstein O. Treating patients not numbers: the benefit and burden of lowering TSH newborn screening cut-offs. *Arch Dis Child*. 2011;96(2):121–122
36. Rapaport R. Thyroid function in the very low birth weight newborn: rescreen or reevaluate? *J Pediatr*. 2002;140(3):287–289
37. O'Grady MJ, Cody D. Subclinical hypothyroidism in childhood. *Arch Dis Child*. 2011;96(3):280–284
38. Alm J, Hagenfeldt L, Larsson A, Lundberg K. Incidence of congenital hypothyroidism: retrospective study of neonatal laboratory screening versus clinical symptoms as indicators leading to diagnosis. *Br Med J (Clin Res Ed)*. 1984;289(6453):1171–1175
39. Azizi F, Afkhami M, Sarshar A, Nafarabadi M. Effects of transient neonatal hyperthyrotropinemia on intellectual quotient and psychomotor performance. *Int J Vitam Nutr Res*. 2001;71(1):70–73
40. Leonardi D, Polizzotti N, Carta A, et al. Longitudinal study of thyroid function in children with mild hyperthyrotropinemia at neonatal screening for congenital hypothyroidism. *J Clin Endocrinol Metab*. 2008;93(7):2679–2685
41. Tyfield LA, Abusrewil SSA, Jones SR, Savage DCL. Persistent hyperthyrotropinaemia since the neonatal period in clinically euthyroid children. *Eur J Pediatr*. 1991;150(5):308–309

42. Miki K, Nose O, Miyai K, Yabuuchi H, Harada T. Transient infantile hyperthyrotrophinaemia. *Arch Dis Child*. 1989;64(8):1177–1182
43. Cody D, Kumar Y, Ng SM, Didi M, Smith C. The differing outcomes of hyperthyrotropinaemia. *J Pediatr Endocrinol Metab*. 2003;16(3):375–378
44. Köhler B, Schnabel D, Biebertmann H, Gruters A. Transient congenital hypothyroidism and hyperthyrotropinemia: normal thyroid function and physical development at the ages of 6–14 years. *J Clin Endocrinol Metab*. 1996;81(4):1563–1567
45. Zung A, Tenenbaum-Rakover Y, Barkan S, et al. Neonatal hyperthyrotropinemia: population characteristics, diagnosis, management and outcome after cessation of therapy. *Clin Endocrinol (Oxf)*. 2010;72(2):264–271
46. Demirel F, Bideci A, Çamurdan MO, Cinaz P. L-thyroxin treatment in infants with hyperthyrotropinaemia: 4-year experience. *Int J Clin Pract*. 2007;61(8):1333–1336
47. Reuss ML, Paneth N, Pinto-Martin JA, Lorenz JM, Susser M. The relation of transient hypothyroxinemia in preterm infants to neurologic development at two years of age. *N Engl J Med*. 1996;334(13):821–827
48. Meijer WJ, Verloove-Vanhorick SP, Brand R, van den Brande JL. Transient hypothyroxinaemia associated with developmental delay in very preterm infants. *Arch Dis Child*. 1992;67(7):944–947
49. La Gamma EF, van Wassenae AG, Ares S, et al. Phase 1 trial of 4 thyroid hormone regimens for transient hypothyroxinemia in neonates of <28 weeks' gestation. *Pediatrics*. 2009;124(2):e258–e268
50. Ng SM, Turner MA, Gamble C, Didi M, Victor S, Weindling AM. TIPIT: A randomised controlled trial of thyroxine in preterm infants under 28 weeks' gestation. *Trials*. 2008;9:17
51. McManus V, Guillem P, Surman G, Cans C. SCPE work, standardization and definition—an overview of the activities of SCPE: a collaboration of European CP Registers. *Zhongguo Dang Dai Er Ke Za Zhi* 2006;8(4): 261–265