Finding Blunders in Thyroid Testing: Experience in Newborns

Laurel J. Zilka,¹ John A. Lott,^{2*} Lula C. Baker,¹ and Sharon M. Linard¹

¹The Ohio Department of Health Newborn Screening Laboratory, Reynoldsburg, Ohio

²Department of Pathology, O.S.U. Medical Center, Columbus, Ohio

We evaluated thyroxin (T4) and thyroid-stimulating hormone (TSH) data along with clinical information from 600,000 newborns. We looked for certain combinations of tests and clinical data that were questionable and possibly mistaken. Our approach suggests that certain combinations of test results, especially the presence of missing results, deserved further evaluation for possible blunders. We found that missing tests were frequently the result of oversight. The laboratory used the well-known standard blood-spot-on-filter paper methods for TSH and T4. For quantitation of TSH and T4, we used the time-resolved fluoroimmunoassay available from Perkin Elmer. We found 56 babies with confirmed primary congenital hypothyroidism (PCH) in a total of 600,000 patients. We also found 18 sets of results in the same 600,000 babies that gave inconsistent findings, had missing values, and (or) possible misinterpretations of the clinical and (or) laboratory data. What

is an acceptable mistake rate? All mistakes are unacceptable, but there is likely some irreducible mistake rate, and efforts to reduce the mistake or blunder rate still further may not be cost-effective. What can be done is to study the mistake rate per 600,000 babies from year to year; the mistake rate should be decreasing or not changing. This assumes a stable cohort of babies; an assumption that may be acceptable. We applied a form of pattern recognition to identify cases of possible blunders and missing values in either the laboratory or clinical data. What is clear is that we apparently identified some blunders. The 18 mistakes per 600,000 babies may be “very low” and acceptable. We recommend that seeking ever decreasing mistakes is the way to go, and the level of monitoring the data should be very intense given the serious consequences of misdiagnosed thyroid disorders. *J. Clin. Lab. Anal.* 22:254–256, 2008. © 2008 Wiley-Liss, Inc.

Key words: neonatal screening; hypothyroidism; laboratory blunders; laboratory mistakes; primary congenital hypothyroidism (PCH)

INTRODUCTION

An ongoing area of vital interest in clinical laboratories is the relationship between laboratory and (or) clinical information and possible bogus values. The usual assumption is that the clinical or diagnostic information is more subjective, i.e., “softer” than the hard laboratory numbers. Laboratory data are, of course, not absolute, and they are affected by a host of variables; some of these are described elsewhere (1). Diagnostic data may be confounded by a collection of variables that change their use and importance. The most serious problem is missing data. As part of our ongoing QC efforts, we chose to evaluate both the thyroid-stimulating hormone test (TSH, thyrotropin) and the thyroxin test (T4) in newborn babies. We included babies based on their screening test results where there was a suspicion of primary congenital

hypothyroidism (PCH). We included those patients who had “moderate-” or “high-” risk thyroid findings. The flow diagram used by us to define risk is elsewhere (2).

MATERIALS AND METHODS

Patients (Newborns)

A total of 600,000 newborns were tested for TSH and T4. Follow-up diagnostic information was obtained for

*Correspondence to: John A. Lott, Department of Pathology, O.S.U. Medical Center, 1645 Neil Avenue, Columbus, OH 43210.
E-mail: lott.1@osu.edu

Received 28 February 2008; Accepted 5 March 2008

DOI 10.1002/jcla.20247

Published online in Wiley InterScience (www.interscience.wiley.com).

TABLE 1. Possible Blunder(s) in Neonatal Thyroid Testing

Change of diagnosis	Number of subjects (babies). Total = 18	Reasons for change in diagnosis
1 YES to NO	5	Normal results were obtained from a thyroid challenge test. Thyroxin was not restarted
2 YES to NO	3	Hypothyroidism was secondary to maternal Graves Disease. Presence of isovaleric acid acidemia or thyroid dysgenesis. Presence of ectopic thyroid gland
3 Unknown to NO	1	Normal results were obtained from a thyroid challenge test and thyroxin was not reinstated
4 Unknown to NO	1	Follow up testing indicated a normal baby and thyroxin was not given
5 NO to YES	1	Physician listed diagnosis as compensated hypothyroidism, indicating adequate compensation with thyroxin
6 Unknown to YES	1	Thyroid challenge was performed and elevated TSH concentrations were obtained on follow-up testing
7 Unknown to YES	1	Child continues to receive thyroxin
8 Unknown to YES	2	Follow-up laboratory testing indicating PCH
9 Unknown to YES	1	Thyroxin was restarted after elevated TSH concentrations were obtained following a thyroid challenge
10 Unknown to YES	1	Child, 3 yr, was last seen by the primary-care physician. Follow-up laboratory testing until that time was consistent with PCH
11 YES to Unknown	1	Child was last seen by primary-care physician when 3 yr old. Follow-up laboratory results were not obtained

TABLE 2. The Physicians' Diagnoses were Partly Unknown for the Reasons Listed Here and Include Parental Noncompliance and Insufficient Follow-up

Number of babies	Reason for diagnosis of "unknown"
2 (twins)	Mother noncompliant
2	Mother never followed up with primary-care physician or specialist
1	Baby's current primary-care physician was not physician of record at the time and received no notifications
1	Baby's specialist did not share information with the primary-care physician

the "moderate-" or "high"- risk babies to compare the laboratory- and physician-derived diagnostic findings.

Methods

We used a time-resolved fluoroimmunoassay method for both TSH and T4 on filter-paper-punched specimens with the Auto-DELPHIA instrument from Perkin Elmer (3–5). Based on our earlier results (2), we used a $\geq 20 \mu\text{IU/ml}$ cutoff for TSH to identify babies of any age who were at "moderate risk" or at "high risk" and who then required confirmatory testing.

SCREENING LABORATORY RESULTS

Incidence of Risk

Out of our cohort of 600,000 babies, and using our algorithm, we found 165 newborns who were at

"moderate" or "high" risk for PCH. Of these, we found 92 babies at "high" risk and 73 babies who were at "moderate risk".

Physicians' Diagnoses

Of the 92 babies with confirmed, high-risk positive screens, 56 were diagnosed to be positive for PCH. In 30 of these, the physicians' diagnoses were of "low" risk. As a percentage of all findings, one physician stood out with consistently more cases of PCH per 100 babies as compared with other physicians. The reason for this is unclear.

Change of Diagnosis

Table 1 shows instances for 18 babies where a change in the diagnosis occurred indicating a possible blunder. We believe that a blunder has occurred in most cases when a change in diagnosis is present. These changes could be used to trap potentially incorrect results. We give here 11 patterns that must be reviewed for possible blunders. In the table, YES always means disease (PCH) is present and NO means that disease (PCH) is absent; Unknown indicates that the diagnosis is unknown (see Table 2). The most common blunder was simply missing values for any of the tests.

A change in the diagnosis comes from the data in column 3. If the change is incompatible with the data in column 1, a blunder has occurred. For example, a change of YES to NO in row 1 of the table means that after a review of the additional thyroid challenge test, disease is diagnosed as absent; i.e., the baby is normal. If

the additional test is not performed, the diagnosis is incorrect and a blunder has occurred. As another example if in the 11th row (the last row), the diagnosis goes from YES to Unknown because key information (follow-up laboratory results) was missing, a blunder has occurred. Review of all the data and (or) follow-up tests changes the diagnosis. All the other cases in the table were analyzed the same way. In general, a change of diagnosis warrants further evaluation of that patient to look for bogus or missing data. Of course, not all possible combinations of laboratory and physician's responses are here. Laboratories have to create combinations of these findings to identify rogue or bogus values in PCH.

CONCLUSIONS

The patterns in the table could serve as blunder-detection algorithms. Most blunders arise owing to missing data or diagnostic conclusions from incomplete data. All actors in the laboratory testing and diagnosis play a role in blunder detection. The clinical information is particularly important, and we found it is often difficult to obtain. There are of course limitations to our approach. The blood spot tests for TSH and T4 should not be over-interpreted. They are not diagnostic tests, and confirmatory tests, such as quantitative tests for TSH and T4, are needed to make a diagnosis. A challenge test that measures T4 during a period of T4 deprivation should be employed in any baby at "moderate" or "high" risk. It is only when the screening and quantitative tests for TSH and T4 and possibly challenge results are on hand that a diagnosis can be

approached. The clinical findings must of course be available. Our approach is to first identify babies in the "moderate-" or "high"-risk categories and then superimpose the clinical data to see whether there is a real change in diagnosis, if diagnostic tests are absent, or if there has been a bogus shift of the data.

The mistake rate in our study is low (18 out of 600,000 cases), and we would have very likely missed some babies given the structure of our study. Other algorithms could be constructed, and they could increase the sensitivity of the blunder detection. The use of incomplete data to make a diagnosis is the major source of blunders. We make no recommendation as to an acceptable mistake rate. There is really no satisfactory mistake rate; what can be said is that the mistake rate should always be decreasing. The basis of all mistake detection is constant vigilance. With further information, likely patterns of results could be developed for enhanced blunder detection.

REFERENCES

1. Grannis GF, Gruemer H-DF, Lott JA, et al. Proficiency evaluation of clinical chemistry laboratories. *Clin Chem* 1972;18:222-236.
2. Lott JA, Sardovia-Iyer M, Speakman KS, et al. Age-dependent cutoff values in screening newborns for hypothyroidism. *Clin Biochem* 2004;37:791-797.
3. Perkin-Elmer Life Sciences, Tuku, Finland.
4. Miyai K, Miyage T, Ashida N, et al. Evaluation of highly sensitive thyrotropin assay for detecting thyroid diseases in neonatal screening: Preliminary studies. *Endocr J* 1998;45:761-766.
5. Mei JV, Alexander JR, Adam BW, et al. Use of filter paper for the collection and analysis of human whole blood specimens. *J Nutr* 2001;131:1631S-1636S.