Assessment of Parental Understanding by Pediatric Residents During Counseling After Newborn Genetic Screening

Michael H. Farrell, MD; Pramita Kuruvilla, MD

Objective: To investigate pediatric residents’ efforts to assess understanding in discussions about positive newborn screening test results. Newborn screening saves lives, but confusion about false-positive and carrier results often leads to psychosocial problems.

Design: Explicit-criteria abstraction of transcripts of encounters with standardized parents of a fictitious infant found to carry cystic fibrosis or sickle cell hemoglobinopathy.

Setting: Simulated doctor-patient encounter.

Participants: Pediatric residents participating in an educational workshop on how to inform parents about positive newborn screening test results.

Main Outcome Measures: Abstraction used an explicit-criteria data dictionary with definitions for 5 different ways to assess understanding. A “partial” designation was used for leading syntax or no pause for response.

Results: Interabstractor reliability over 59 transcripts (2 per resident) was $\kappa = 0.93$. Only 26 of 59 transcripts (44.1%) met definite criteria for at least 1 assessment of understanding. Most assessments were the less effective close-ended (37.3% of transcripts) and “OK?” question types (32.2% of transcripts). Only 3 transcripts met definite criteria for an open-ended assessment and no transcripts included a request for a teach-back, the type thought to be most effective. Four transcripts (6.8%) included an advance request for questions. With partial-criteria assessments included, an additional 31 transcripts (52%) were identified.

Conclusions: The small number of assessments of understanding and the high fraction of less effective assessments do not bode well for parental understanding, especially for parents with limited health literacy. Training programs should address assessments of understanding, but quality improvement activities using these types of assessment methods may also be needed.

Arch Pediatr Adolesc Med. 2008;162(3):199-204

Communication is said to be the “main ingredient” of medical care and “the vehicle by which technical care is implemented and on which its success depends.” Unfortunately, communication problems have been documented across much of health care. Medical schools and residency programs provide training for communication skills, but without periodic reinforcement, it is unclear whether the benefits of training will persist for long. It also seems unlikely that physicians already in practice will have time or resources necessary for either training or refresher courses. We have therefore been working to adapt techniques from quality improvement to the needs of communication. Quality improvement has a track record for changing physician behavior on a population scale and uses fewer resources than are necessary for educational programs. We use quality indicators to operationalize individual communication behaviors. Quality indicators are explicitly defined, quantitatively reliable measures of ideal clinical behaviors that each represent a small domain within the overall quality of care. In this study, we used a quality indicator method to analyze pediatric residents’ efforts to assess parent understanding during a discussion about a newborn screening result.

Assessment of understanding questions (AU) (Table 1) is important for those many situations when a physician needs to explicitly verify a patient’s knowledge about a topic or that a message has been successfully conveyed. Assessment of understanding questions is espe-
families are asked to adhere to a complicated treatment plan, and patients who might not be displaying their feelings because of a sense of shock or for cultural reasons. Physicians can also include a statement early in each encounter encouraging the patient to interrupt if questions arise (an “advance request for questions”).

We are especially concerned about use of AUs after newborn screening. Nearly every US newborn is screened for a panel of life-threatening diseases, but most “abnormal” results turn out to be false positives or reveal that the infant is a genetic carrier for cystic fibrosis (CF) or sickle cell hemoglobinopathy (SCH). Some authors have referred to these infants as having a “nondisease” since they do not become sick but their parents can develop psychosocial complications such as anxiety, depression, or stigmatization or misconceptions about whether the infant has a disease. Confusion could be partially due to primary care physicians’ role in informing parents, since communication in primary care has been criticized by families and public health officials. Newborn screening programs provide educational materials to parents and physicians, but in a recent survey, we found that most programs lack a follow-up mechanism to monitor psychosocial outcomes. Families of carrier infants can be helped by genetic counselors, but the supply of counselors is limited enough that most families will not be able to access their services.

These communication failings and psychological complications are often cited by ethicists and policy experts in arguments against the routine use of genetic and molecular screening technologies. To keep newborn screening viable, we have argued that the screening programs or referral centers should introduce population-scale programs to assess and improve the quality of communication services and parents’ psychological outcomes. These types of programs may ensure that newborn screening results in more good than harm, similar to the way that biomedical follow-up mechanisms ensure that infants with diseases receive appropriate care and have favorable outcomes. Once developed for newborn screening, population-scale communication quality-assurance techniques may be applicable to many other communication problems in health care.

### Table 1. Types of AU

<table>
<thead>
<tr>
<th>AU Type and Definition (Relative Effectiveness Shown in Parentheses)</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Request for teach-back (most effective): A question whose natural answer is the patient’s own version of the information covered in conversation. Some clinicians include an explanation of why they are requesting the teach-back.</td>
<td>To make sure that I was being clear enough, could you please share with me the main points you got from our discussion?</td>
</tr>
<tr>
<td>Open-ended AU (effective): An assessment question that places few or no restrictions on response. In many cases, the natural answer will be an actual question.</td>
<td>You may have to discuss this issue with your wife, so let’s practice now how you will explain this to her.</td>
</tr>
<tr>
<td>Close-ended AU (less effective): Assessment to which the natural answer is restricted, such as yes, no, or some other brief, limited answer.</td>
<td>What questions do you have? What parts of this are hard to understand?</td>
</tr>
<tr>
<td>“OK?” question (ambiguous; could be interpreted in many ways): Close-ended assessment that does not specifically mention understanding or a question. The patient may interpret this assessment type as a question about emotions or as a request for permission to proceed.</td>
<td>Do you have any questions? Does that make sense? OK? All right?</td>
</tr>
</tbody>
</table>

Abbreviation: AU, assessment of understanding questions.

### METHODS

#### DESIGN

To determine the frequency and type of AUs, we developed a new explicit-criteria procedure to abstract transcripts of conversations between pediatric residents and standardized parents of a fictitious infant whose newborn screening test results suggested carrier status for CF or SCH. Methods were approved by institutional review boards at Yale and the Medical College of Wisconsin.

#### PARTICIPANTS AND SETTINGS

The participants were all residents beyond their first year in a prominent pediatric residency program. Residents were scheduled in groups of 4 to attend a workshop entitled “Communicating Genetic Test Results: Newborn Screening for Cystic Fibrosis and Sickle Cell Hemoglobinopathy.” The residents were told that the objectives of the workshop included to “learn about newborn screening” and “practice your communication with 2 simulated patients.” Each workshop began with a 10-minute review of newborn screening, CF, SCH, and autosomal recessive inheritance. The review avoided any discussion of how to discuss results. The workshops were part of the official curriculum, but residents gave informed consent and were free to decline the use of their tapes in research.

#### PROCEDURES

**Counseling Task**

Each resident was audiotaped in 2 separate encounters. A handout described the screening result and some background data but did not prompt how to inform the parent. In the SCH carrier scenarios, the handout reported a screening result of hemoglobin F, A, and S, a result that had been presented in the review session as definitely indicating that an infant is an SCH carrier. In the “likely CF carrier” scenarios, the handout reported an elevated screening immunoreactive trypsinogen level and the presence of 1 ΔF508 mutation with no multiallele follow-up screening. The review sessions had presented such a result as suggesting that the infant was probably a carrier but still had a 5% to 10% chance of having the disease because of an undetected allele. The infant’s mother and father were por-
trayed in the scenario as both being adopted so that the session could focus on risk communication rather than on taking a family history.

During the year, 6 female standardized parents worked on the project; each was chosen to plausibly depict the age and ethnicity of a mother of an infant with CF or SCH. The order of encounters for each resident was randomly determined in double-crossover fashion over the 2 diseases and 2 script types. The 2 script types were included for a separate validation study of the new script type, the Brief Standardized Communication Assessment. The Brief Standardized Communication Assessment is a streamlined version of a standardized patient script that requires less training, discourages improvisation, and encourages counseling to continue via nonleading questions like “Is there anything else I should know?” The comparison script resembled scripts used in most educational settings, which seek to make the encounter realistic by asking the patient to use acting skills and improvise a 2-way dialogue. All standardized parents were coached not to appear anxious, to avoid requests for clarification, and to adopt a “polite nod” body language that does not necessarily denote understanding. These instructions helped our analysis to focus on AUs rather than the resident’s ability to reclarify for parents who appear mystified (an important skill but not the subject of this study).

Data Handling

We taped 32 residents in 2 interviews apiece, although taping equipment failed during part of 5 interviews so the final sample consisted of the remaining 59 transcripts. Tapes were transcribed verbatim and proofread for accuracy by a board-certified pediatrician. We used a sentence diagramming procedure to divide transcripts into individual “statements,” each with 1 subject and 1 predicate.

Abstraction

We adapted the transcript abstraction procedure from methods used in medical record review. Abductors were guided by a quality improvement–style data dictionary derived from existing guidelines. The data dictionary contained explicit-criteria definitions and examples for 4 different types of AUs (Table 1) and for the advance request for questions. For a “definite” designation, the AU had to be followed by a pause for parent response, even for a simple continue like “uh huh.” A “partial” AU designation was used for AUs that were not followed by a pause or for AUs that used a leading syntax (eg, “You don’t have any questions, do you?”). The “OK?” question AU was only allowed for instances when the word “OK” was transcribed as an interrogative because of a raising of voice pitch.

Half of the transcripts were duplicatively abstracted to assess interabstractor reliability, and following the suggestion by Feinstein, half of the duplicates were discussed afterward to ensure quality control and consistency. Discrepancies between abstractors were automatically resolved by a spreadsheet to avoid subjective judgment.

ANALYSES

Statistical Analysis

Statistical analysis was done using JMP software (SAS Institute, Cary, NC). Interabstractor reliability was calculated using the Cohen method with an adaptation for ordinal agreement. Variables, such as the presence of the various AU types, were analyzed using 1-way analysis of variance for continuous responses for categorical variables and the χ² test for grouped categorical responses.

RESULTS

Descriptive data on the participants (Table 2) were similar to those of the population of the residency program at the time of the study. Interviews ranged from 5 to 20 minutes (mean, 9.8 minutes). A total of 144 full AUs and 551 partial AUs were identified in 102 135 words of transcribed conversation. Interabstractor reliability corrected for chance was κ = 0.93, with separate coefficients for AU types of 1.0 for advance request for questions, 0.33 for open-ended questions (which were rare), 0.85 for close-ended AUs, and 0.95 for “OK?” questions.

NUMER AND TYPES OF AUs INCLUDED

Only 26 of 59 transcripts (44.1%) met definite criteria for at least 1 definite AU (Table 3). Among these, the mean (SD) number of AUs per transcript was 5.5 (7.6). Definite criteria for an advance request for questions were found in 4 of 59 transcripts (6.8%). There were no statistically significant associations between AU number and the resident’s sex or year in residency, duration of counseling, or a screening result of CF or

**Table 2. Participant Characteristics**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>No. (%) Responding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>9 (29)</td>
</tr>
<tr>
<td>F</td>
<td>22 (71)</td>
</tr>
<tr>
<td>Age, y¹</td>
<td></td>
</tr>
<tr>
<td>25-29</td>
<td>22 (73)</td>
</tr>
<tr>
<td>30-34</td>
<td>8 (27)</td>
</tr>
<tr>
<td>Year in residency</td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>19 (61)</td>
</tr>
<tr>
<td>3rd or 4thb²</td>
<td>12 (39)</td>
</tr>
</tbody>
</table>

¹Thirty residents responded to the age question.
²Fourth-year residents were from the medicine-pediatrics program.

Letter Grades

Quality improvement interventions often involve feedback to a targeted subgroup most in need of improvement, so as with our earlier studies, this analysis included a procedure for adapting data for clinician feedback. The problem with feedback about AU questions is that there is no AU taxonomy that would be familiar to most clinicians. To help categorize AU quality indicator performance for future interventions, scores were grouped ordinarily using the familiar letter grades of A through F, with C and lower grades denoting clinicians in need of improvement. Comparisons across transcripts were made using a grade point average system of zero to 4 points. Given the recent recommendations for requests for a teach-back in the health literacy literature, a definite-criteria example of this AU type was used to define an A grade. At the other end, an F grade was used for transcripts that contained no AUs of any type and the D grade, for AU attempts such as “OK?” questions or partial-criteria AUs. This left the C grade for close-ended AUs, and the B grades for open-ended AUs.
The percentages of transcripts meeting definite criteria for each AU type are shown in Table 3. Only a few transcripts contained definite criteria for an open-ended AU, and no transcripts contained definite criteria for a request for a teach-back. The less effective close-ended and “OK?” question AUs were far more common. Fortunately, the “OK?” question was the sole type of AU in only 3 of 26 transcripts with AUs (11.5%).

### EFFECT OF USING PARTIAL ABSTRACTION CRITERIA

As mentioned earlier, the definite/partial/absent scheme helped to differentiate between AUs and statements that may have been intended as AUs but were less likely to be effective. When abstractors’ partial ratings were considered, at least 1 AU “attempt” was seen in an additional 31 transcripts (52.5%), so that 57 of 59 transcripts included either definite or partial criteria for at least 1 AU. The overall mean (SD) number of AUs increased with partial criteria to 12 (14.9) per transcript, but there were still no statistically significant associations with case or resident characteristics. The AU type most responsible for the increase with partial criteria added was the “OK?” question, which was seen with no pause for response in 21 of the 31 new transcripts (63.6%), with a range of up to 63 instances in a single transcript.

### TRANSCRIPT COMPARISON AND THE HYPOTHETICAL INTERVENTION

Under the letter-grade definitions proposed earlier, the mean (SD) grade point average would have been 1.4 (0.7), a D+ average. Residents were moderately consistent across their 2 transcripts; 22 residents (73%) had the same letter grade for both transcripts, while 6 (20%) differed by only 1 grade. No transcripts would have qualified for an A grade because none of the residents requested a teach-back from either of the standardized parents. Three transcripts (34%) would have received a B grade, 20 transcripts (34%) had criteria for a C grade, and 34 transcripts (58%) had criteria for a D grade. Two transcripts would have been rated with an F grade with no partial or definite AUs at all.

Thus, if transcripts qualified for targeted feedback on the basis of a C or lower grade, then a total of 56 individual transcripts (94.9%) would have received feedback, or all of the residents on at least 1 of their transcripts. This situation is anything but the “targeted” feedback situation preferred in many quality improvement projects.

The purpose of this study was to investigate pediatric residents’ efforts to assess parental understanding during discussions about a positive newborn screening test result. Most transcripts included at least 1 AU question, but most AUs were the close-ended or “OK?” question type. Only 3 residents included an open-ended AU and none of the residents asked for a teach-back, the communication behavior thought to be most effective at assessing understanding.

These findings have important implications for medical education, newborn screening, and health care in general. Physicians are often asked to explain complicated subjects to patients, but unless patient understanding is explicitly verified, the physician may never know if counseling was effective. Unfortunately, the stakes are high; in the case of newborn screening, misunderstandings can lead to anxiety, depression, vulnerable child syndrome, stigmatization, or reproductive or medical decision making that the parents later come to regret. Physicians may be less oblivious to these complications if they can be taught to include AU questions in counseling. Such a focused educational strategy seems more like the often-disparaged “bag-of-tricks” model of communication than like the global paradigm that is most common in educational settings. The difference between focused and global perspectives was also anecdotally evident in the educational workshop aspect of the project, in which taped conversations were evaluated much more favorably than they might have been if preceptors had had access to AU quality indicator scores at the time of evaluation. If such discrepancies between global and specific perspectives are found in future studies, medical schools and residency programs should consider implementing more behavior-specific teaching and evaluation methods.

On a population scale, where a traditional educational approach would require too many resources and too much time from physicians, our methodological approach may finally enable quality improvement professionals to assess and improve communication. To evaluate the feasibility of our methods, we tracked time and expenses and project that quality improvement projects in the future will be achievable with minimal training and costs less than $50 per physician. Use of an explicit-criteria abstraction technique instead of a subjective rating scale should help reduce bias and improve interabstractor reliability and the ability to compare or rank physicians; the κ coefficient of 0.93 is excellent for a social science method at an early stage of development.

This study was limited by its small sample size, but we see some limitations as strengths from a quality improvement perspective. Qualitative methods would have

---

**Table 3. Transcripts With Definite Criteria for at Least 1 AU**

<table>
<thead>
<tr>
<th>AU inclusion</th>
<th>Transcripts, No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No definite AUs</td>
<td>33 (55.9)</td>
</tr>
<tr>
<td>At least 1 AU</td>
<td>26 (44.1)</td>
</tr>
<tr>
<td>AU type</td>
<td></td>
</tr>
<tr>
<td>Request for a teach-back</td>
<td>0</td>
</tr>
<tr>
<td>Open-ended</td>
<td>3 (5.1)</td>
</tr>
<tr>
<td>Close-ended</td>
<td>22 (37.3)</td>
</tr>
<tr>
<td>“OK?” question</td>
<td>19 (32.2)</td>
</tr>
</tbody>
</table>

Abbreviation: See Table 1.

a Some transcripts contained more than 1 type of AU.
provided a richer description of conversation, but qualitative methods have limited reliability and would be prohibitively expensive for use in quality improvement. The use of standardised parent encounters for quality improvement projects instead of real parents avoids logistical, privacy, and consent difficulties associated with audiotaping actual clinic visits. Simulation is also useful because a sense of observation prompts clinicians to greater efforts; the resulting competence data approximate a ceiling for process of care because competence is necessary but not sufficient for performance.  

Simulation also allows an equal-footing comparison across clinicians that would be impossible in analyses of actual encounters. Even so, we reduced artificiality of simulation by holding the workshops in the residents’ continuity clinic. We saw residents as ideal for this demonstration project since many are near the peak of their content knowledge on these topics.

Further study will be necessary to investigate a possible relationship between AU use and communication outcomes, but such studies will need to examine more complicated variables than whether an AU is included anywhere in conversation. We have therefore developed a new “load” calculation to represent patterns of AU placement in conversation (M.H.F. and S. Christopher, MA, unpublished data, 2007). Until more research has been done about AU placement relative to informational content, it is important for all health care providers to at least recognize that guidelines advise assessment of understanding at several points in conversation.

In sum, the quality of communication observed in this study was marked by many missed opportunities to assess understanding, suggesting to us a route by which many parents will develop psychological problems. Communication about newborn screening is challenging because of differences in testing methods, diseases, and families,  

and psychosocial complications of poor communication are in part inhibiting society from implementing new screening programs. If integrated into US newborn screening practices, communication quality assurance may benefit clinicians and families and help satisfy society’s ethical concerns about the risk and benefit with expanded newborn screening and genetic screening. Other communication quality assurance programs can also be developed to improve communication processes and outcomes in every other area of health care.

Accepted for Publication: July 17, 2007.

Correspondence: Michael H. Farrell, MD, Center for Patient Care and Outcomes Research, Medical College of Wisconsin, 8701 Watertown Plank Rd, Milwaukee, WI 53226 (mfarrell@mcw.edu).


Funding/Support: Dr Farrell is supported in part by grant K01HL072530 from the National Heart, Lung, and Blood Institute.

Additional Information: When the project began, Dr Farrell was assistant professor at the Yale Primary Care Medicine Research Center (Waterbury, Connecticut) and Dr Kuruvilla was a medical student in the Yale University School of Medicine (New Haven, Connecticut).

Additional Contributions: Jeffrey Stein, MD, Lynnea Ladouceur, MPH, and Stephanie Christopher, MA, provided valuable advice and assistance.

REFERENCES

25. Makoul G. Essential elements of communication in medical encounters: the Kalama-
217-222.
27. Silverman J, Kurtz S, Draper J. Skills for Communicating With Patients. Abing-
29. Cordero J, Edwards ES, Howell RR, et al. Newborn screening: toward a uniform screening panel and system. A report by the American College of Medical Ge-
netics (ACMG) for the Maternal and Child Health Bureau (MCHB) of the US Health Resources and Services Administration (HRSA). http://mchb.hrsa.gov
31. Hampton ML, Anderson J, Lavizzo B, Bergmen A. Sickle cell "nondisease": a po-
33. Tluczek A, Kosick R, Farrell P, Rock M. Psychosocial risk associated with new-
born screening for cystic fibrosis: parents’ experience while awaiting the sweat-
13(3):181-186.
35. Tluczek A, Mischler E, Bowers B, et al. Psychological impact of false-positive re-
sults when screening for cystic fibrosis. Pediatr Pulmonol Suppl. 1991;7:
29-37.
102(1, pt 1):44-52.
37. Markel H. Scientific advances and social risks: historical perspectives of genetic screening programs for sickle cell disease, Tay-Sachs disease, neural tube de-
38. Billings PR, Kohn M, de Cuevas M, Beckwith J, Alper J, Natowicz M. Discrimi-
476-482.
39. Giske DJ, Haavisto A, Laxova A, Rock L, Farrell P. Genetic counseling and neo-
40. Farrell M, Certain L, Farrell P. Genetic counseling and risk communication ser-
(2):120-126.
42. Bobadilla JL, Farrell MH, Farrell PM. Applying CFTR molecular genetics to facili-
tate the diagnosis of cystic fibrosis through screening. Adv Pediatr. 2002;49:
131-190.
43. Hoppe Ruth B. Standardized (simulated) patients and the medical interview. In:
Lipkin M, Putnam SM, Lazare A, eds. The Medical Interview: Clinical Care, Edu-
45. Feinstein AR. Clinical Epidemiology: The Architecture of Clinical Research. Phila-
47. Farrell MH, Kuruvilla P, Brienza RS. Assessment of understanding: a quality in-
dicator for communication before adult cancer screening. J Gen Intern Med

Thank You to Our Reviewers

“You mean you do that for free? Why would you spend the time? What’s in it for you?”

These might be the comments to our peer reviewers from someone outside the scientific community. Our peer reviewers fit the adage that if you want something done, you ask a busy person. They are people with many other demands who freely give their time and expertise to Archives by carefully reading manuscripts, thinking about them, writing detailed reviews, and sending them to us. These reviews are a key element in our decision on how to proceed with submitted manuscripts. The reviewers’ comments make the manuscript better and thus contribute in an immeasurable way to the scientific process. They help us understand the importance of a piece of work and how it makes a contribution to the field. They identify errors before they are published and make the science of these publications stronger.

The list of individuals who have served as peer reviewers for us this year is published on our Web site at http://www.archpediatrics.com. Our community of authors, readers, Archives, and the editors are in your debt.

Thank you.

Frederick P. Rivara, MD, MPH
Editor