

Methods and Criteria Used in Evidence-Based Decisions in Public Health

Coleen A. Boyle,¹ Paul Ladenson,² and James E. Haddow³

A workshop entitled, "The Impact of Maternal Thyroid Diseases on the Developing Fetus: Implications for Diagnosis, Treatment, and Screening," was held in Atlanta, Georgia, January 12–13, 2004. This paper reports on the session that examined methods and criteria used for decisions in public health. For this session the following papers were presented: "Methods to Evaluate Scientific Evidence," "Criteria for Screening," and "Public Health Considerations." Development of evidence-based guidelines, strengthened by rigorous systematic reviews, will improve the quality, efficiency, and cost effectiveness of management of thyroid dysfunction among reproductive-age women. Maternal and fetal benefits that have been hypothesized to result from screening pregnant and pre-pregnant women for hypothyroidism include reduced incidences of peripartum maternal complications and fetal loss and optimization of fetal and neonatal neuropsychological development. Screening should be considered as the initial step in a comprehensive program that includes appropriate diagnostic and therapeutic interventions. The actual benefits and potential risks (i.e., iatrogenic thyrotoxicosis) of implementing a thyroid function screening program have not been demonstrated in a prospective randomized clinical trial or prospective cohort study. Consequently, it is difficult to develop consensus and secure resources for a comprehensive thyroid function screening and therapeutic intervention program in women who are or anticipate becoming pregnant. Marshalling support for performance of both a clinical trial and high-quality observational studies should be a high priority.

Introduction

THIS REPORT contains the summaries of several papers presented to a workshop held in Atlanta, Georgia, January 12–13, 2004 to address "The Impact of Maternal Thyroid Diseases on the Developing Fetus: Implications for Diagnosis, Treatment, and Screening." The purpose of this session was to describe and discuss the scientific basis for clinical and public health practice. The session was moderated by Dr. Coleen A. Boyle.

"Methods to Evaluate Scientific Evidence" —Dr. Paul Ladenson

There are a number of tools that are used to guide medical practice. These include practice algorithms that outline a stepwise decision tree analysis for clinical practice, clinical practice pathways that guide the sequence of care (often multidisciplinary for a specified patient population), and clinical practice guidelines that are more formal statements to

support practitioner and patient decisions. The Institute of Medicine defines clinical guidelines as "systematically developed statements to assist practitioner and patient decisions about appropriate healthcare for specific clinical circumstances" (1). The aim of such guidelines is to impact the quality, efficacy, and cost effectiveness of health care. Evidence for developing these tools can come from a number of different venues, including clinical surveys (using clinical scenarios to determine usual practice); expert opinion that is often used in areas in which evidence is inadequate; narrative literature reviews; systematic literature reviews and cost-effectiveness studies. Use of expert opinion and narrative reviews, while useful in some situations, are limited in that they have been shown to be biased and lack reproducibility (2,3). The systematic literature review is the most comprehensive and powerful and depends on an assessment of the research quality. The methods used for systematic reviews have been well defined (4,5). Conducting a systematic review, however, is complex and costly. Most practice guide-

¹National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia.

²Division of Endocrinology and Metabolism, The Johns Hopkins University School of Medicine, Baltimore, Maryland.

³Foundation for Blood Research, Scarborough, Maine.

The opinions and findings in this article are those of the authors and should not be construed as official policies or positions of the Centers for Disease Control and Prevention or the U.S. Department of Health and Human Services.

lines in thyroidology rely on narrative literature reviews and expert opinion; few are based on the systematic evaluation of the published literature and few explicitly acknowledge the underlying values of protecting the vulnerable persons and limiting health care expenditures (6).

There are well-defined steps in clinical guideline development. These include defining the clinical questions (a major goal of this workshop); organization of a panel to do the review and assessment; the complex task of reviewing the evidence; and finally the formation of recommendations. A final (although not formal) step is peer review and subsequent revision of the guidelines by not only practitioners, but by patient and advocacy groups, to assure acceptability from a societal perspective.

Considerations related to the development of the guidelines for maternal hypothyroidism include: data on the prevalence of overt and mild hypothyroidism (revealed by elevated thyrotropin [TSH] concentration); data on the consequence of hypothyroidism during gestation for mothers and their exposed children; evidence to support the effectiveness of screening strategies and interventions to treat maternal hypothyroidism; and randomized clinical trials, however they may be appropriate in this area.

Current financing may preclude thyroid screening of women outside of pregnancy. However, when the cost effectiveness of TSH screening was examined per quality-adjusted life year (QALY), it demonstrated comparable (or even better) cost effectiveness to other widely accepted clinical interventions (e.g., screening for breast cancer, hypertension, hypercholesterolemia) (7). Evaluation of cost effectiveness of screening considers not only the costs of screening but risk and benefits associated with screening and subsequent treatment (8,9).

The development of evidence-based guidelines, strengthened by rigorous systematic reviews will improve the quality, efficiency, and cost effectiveness of managing thyroid disorders among reproductive-age women.

In discussion: it was noted that it might not be ethical to conduct randomized clinical trials in this area. Dr. Ladenson acknowledged there may be ethical dilemmas and noted that carefully done epidemiologic studies could provide suitable evidence for clinical decision-making. Another workshop participant cautioned about the potential risks or harm in implementing a new intervention if the benefits are not well justified and the risk clearly known. Dr. Ladenson concurred.

“Criteria for Screening”—Dr. James Haddow

Screening is the systematic application of a test or inquiry to identify subjects at sufficient risk for a specific disorder to benefit from further investigation or direct preventive action, among persons who have not sought medical attention because of symptoms of that disorder (10). Medical screening contains three elements: (1) it is a process of selection with the purpose of identifying those individuals who are at a sufficiently high risk of a specific disorder to warrant further investigation or sometimes direct preventive action; (2) it is systematically offered to a population of people who have not sought medical attention because of the symptoms of the disease for which screening is being conducted; and (3) its purpose is to benefit the individuals being screened.

Wald and Cuckle (11) examined the criteria to be satisfied

before introducing a population screening test. These include the following:

- A well-defined disorder that is sufficiently serious in terms of prevalence, morbidity and mortality to warrant testing;
- Well-defined test attributes including the test to be used, and whether the testing process is centralized;
- Well-defined follow-up diagnostic testing and intervention, including the efficacy of the intervention;
- Well-defined test performance, including an understanding of the distribution of test measurements in both affected and unaffected groups within the population. Test performance includes: the detection rate (or sensitivity), the false-positive rate (or 1-specificity), and the odds of being affected given a positive result (or positive predictive value);
- Medical benefits that outweigh the risks;
- Financial benefits that justify the costs; and
- Known practical implications of implementation, such as whether special facilities are required.

Data from our 1999 study are used in the following example to address some of the above screening criteria. In that study, TSH values were measured in serum samples collected during the second trimester of pregnancy and stored in the freezer, from a cohort of 25,000 women (12). The TSH measurements were performed approximately 8 years later. The data are presented here not to support or refute the idea of screening for thyroid disease, but rather to illustrate how the attributes of screening enumerated above can be evaluated in the context of screening for maternal hypothyroidism during pregnancy. In this example, the disorder is defined as clinically apparent hypothyroidism in women between the time of TSH testing in pregnancy and 10 years later. In our study, approximately 80% of the pregnant women with TSH levels at or above the 98th percentile were not known to be thyroid deficient at the time of testing; 64% of those women were subsequently diagnosed with clinical hypothyroidism. The average time to diagnosis was 5 years. In comparison, only 4% of the control women (TSH < 98th percentile) developed clinical disease. We obtained follow-up TSH measurements 10 years later in the case and control women to confirm thyroid status. Based on these data, the overall prevalence of clinical hypothyroidism in this population was estimated to be 5% over a 10-year period. Although this rate is based on only one study and may not be generalizable, it does illustrate that maternal hypothyroidism may be a very common disorder.

Based on the data presented above, the detection rate for TSH testing during pregnancy is 25% (the proportion of pregnant women identified by the baseline screening test—TSH > 98th percentile or more—who will have clinical hypothyroidism in the next 10 years); the false-positive rate is 0.75% (meaning that 7 or 8 screened women per 1000 might be given thyroid replacement who are not going to go on to develop clinical hypothyroidism), and the odds of being affected given a positive result is 2 to 1 (meaning that two thirds of the women with TSH values 98th or more percentile will subsequently have clinical disease).

Thyroid screening may be easier to implement than some existing prenatal screening processes (such as screening for

Down's syndrome or cystic fibrosis). Scientific data are still needed to guide important considerations, such as cutoff levels for positive screening values and recommended actions. Regarding the health risks and benefits of screening—these will be considered in detail in subsequent sessions in this workshop. Costs of testing and follow up also need to be examined—cost of testing will be lower, if screening is routine in pregnancy.

Discussion

Dr. Cordero asked Dr. Haddow to expand on the issue of a need for centralization of the screening process. Dr. Haddow said the laboratory quality control issues need to be developed more fully and that clinical interpretation and actions need to be standardized to assure uniformity in broad scale implementation. Another workshop participant commented that the issue of screening is much more complex than presented by Dr. Haddow's talk (i.e., maternal thyroid levels change in pregnancy, and there may be other complementary factors to measure) antibody status. Dr. Haddow responded that screening needs to be kept simple and that it was quite remarkable how predictive an elevated TSH value was of future disease in this population.

"Public Health Considerations"—Dr. Coleen Boyle

There is great potential to do benefit from population screening, however, implementation of screening should only be done when we are well aware of the risks and benefits of screening and when effective treatment is in place. The decision to screen on a population level is a balance between the health risks and benefits attributable to early recognition of disease through screening. Screening should not be considered an end in and of itself, but only as the initial step in a comprehensive program that includes appropriate diagnostic and therapeutic intervention.

There are many factors that will be discussed in this workshop that highlight the complexities of whether population screening for maternal thyroid dysfunction is justified on a scientific level. Importantly, workshop presenters will consider the evidence of maternal and fetal benefit from screening—and also the risks associated with such action. We also consider the ethical, legal, and social implications of screening—obligations that suggest that screening is implemented only when we fully understand the consequences of treatment.

References

1. Institute of Medicine 1990 Definition of key terms. In: Field MJ, Lohr KN, (eds) Publications of the Institute of

- Medicine. National Academy Press, Washington, D.C., pp. 33–51.
2. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC 1992 A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *JAMA* **268**:240–248.
3. Khan S, William P, Sutton R 1998 Meta-analyses and large randomized, controlled trials. *N Engl J Med* **338**:59–62.
4. McAlister FA, Clark HD, van Walraven C, Straus SE, Lawson FM, Moher D, Mulrow CD 1999 The medical review article revisited: has the science improved? *Ann Intern Med* **131**:947–951.
5. Shekelle PG, Woolf SH, Eccles M, Grimshaw J 1999 Clinical guidelines: developing guidelines. *BMJ* **318**:593–596.
6. Belin RM, Ladenson PW, Robinson KA, Powe NR 2002 Development and use of evidence-based clinical practice guidelines for thyroid disease. *Endocrinol Metab Clin North Am* **31**:795–817.
7. Danese MD, Powe NR, Sawin CT, Ladenson PW 1996 Screening for mild thyroid failure at the periodic health examination: A decision and cost-effectiveness analysis. *JAMA* **276**:285–292.
8. Eccles M, Mason J, Freemantle N 2000 Developing valid cost effectiveness guidelines: A methodological report from the North of England evidence based guideline development project. *Qual Health Care* **9**:127–132.
9. Russell LB, Siegel JE, Daniels N, Gold MR, Luce BR, Mandelblatt JS. 1996 Cost-effectiveness analysis as a guide to resource allocation in health: roles and limitations. In: Gold M, Seigel J, Russell L (eds) *Cost-Effectiveness in Health and Medicine*. Oxford University Press, New York, pp. 3–24.
10. Wald NJ 1994 Guidance on terminology. *J Med Screen* **1**:76.
11. Wald N, Cuckle H 1989 Reporting the assessment of screening and diagnostic tests. *Br J Obstet Gynaecol* **96**:389–396.
12. Haddow JE, Palomaki GE, Allan WC, Williams JR, Knight GJ, Gagnon J, O'Heir CE, Mitchell ML, Hermos RJ, Waisbren SE, Faix JD, Klein RZ 1999 Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med* **341**:549–555.

Address reprint requests to:
 Coleen A. Boyle, Ph.D.
 National Center on Birth Defects
 and Developmental Disabilities
 Centers for Disease Control and Prevention
 1600 Clifton Road
 Mailstop E-87
 Atlanta, GA 30333

E-mail: cboyle@cdc.gov