

Effective Health Care Program

Future Research Needs Paper
Number 11

Future Research Needs for Diagnosis of Obstructive Sleep Apnea



Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Future Research Needs Paper

Number 11

Future Research Needs for Diagnosis of Obstructive Sleep Apnea

**Identification of Future Research Needs
From Comparative Effectiveness Review No. 32**

Prepared for:

Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2007-10055 I

Prepared by:

Tufts Evidence-based Practice Center, Tufts Medical Center
Boston, MA

Investigators:

Ethan M. Balk, M.D., M.P.H., *Co-Project Lead*
Mei Chung, Ph.D., M.P.H., *Co-Project Lead*
Denish Moorthy, M.B.B.S., M.S., *Research Associate*
Jeffrey A. Chan, B.S., *Research Associate*
Kamal Patel, M.P.H., M.B.A., *Research Associate*
Thomas W. Concannon, Ph.D., *Investigator*
Sara J. Ratichek, M.A., *Project Coordinator*
Lina Kong Win Chang, B.S., *Research Assistant*

**AHRQ Publication No. 12-EHC031-EF
February 2012**

This report is based on research conducted by the Tufts Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10055 I). The findings and conclusions in this document are those of the author(s), who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care researchers and funders of research make well-informed decisions in designing and funding research and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of scientific judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical research and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances.

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials noted, for which further reproduction is prohibited without the specific permission of copyright holders.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

None of the investigators has any affiliation or financial involvement that conflicts with the material presented in this report.

Suggested Citation: Balk EM, Chung M, Moorthy D, Chan JA, Patel K, Concannon TW, Ratichek SJ, Chang LKW. Future Research Needs for Diagnosis of Obstructive Sleep Apnea. Future Research Needs Paper No. 11. (Prepared by the Tufts Evidence-based Practice Center under Contract No. 290-2007-10055 I.) AHRQ Publication No. 12-EHC031-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2012.
www.effectivehealthcare.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

An important part of evidence reports is to not only synthesize the evidence, but also to identify the gaps in evidence that limited the ability to answer the systematic review questions.

AHRQ supports EPCs to work with various stakeholders to identify and prioritize the future research that is needed by decisionmakers. This information is provided for researchers and funders of research in these Future Research Needs papers. These papers are made available for public comment and use and may be revised.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The evidence reports undergo public comment prior to their release as a final report.

We welcome comments on this Future Research Needs document. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Elisabeth U. Kato, M.D., M.R.P.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgements

Stakeholder Panel

Elise Berliner, Ph.D.
Agency for Healthcare Research and Quality
Rockville, MD

Apostolos P. Dallas, M.D.
Virginia Tech-Carilion School of Medicine
and Research Institute
Roanoke, VA

Carolyn M. D'Ambrosio, M.D., FCCP,
DABSM
Director, The Center for Sleep Medicine
Associate Professor of Medicine
Tufts University School of Medicine
Tufts Medical Center
Boston, MA

Daniel J. Gottlieb, M.D., M.P.H.
Director, Sleep Disorders Center, VA
Boston Healthcare System
Associate Professor of Medicine, Boston
University School of Medicine
Boston, MA

Edward Grandi
Executive Director
American Sleep Apnea Association
Washington, DC

Christian Guilleminault M.D., D.Biol.
Sleep Medicine Division, Stanford
University
Stanford, CA

Joseph A. Hanak, M.D.
Chief, Adult Physical Medicine and
Rehabilitation
Tufts Medical Center
Boston, MA

Vishesh Kapur M.D., M.P.H.
Associate Professor of Medicine
University of Washington
Medical Director, UW Medicine Sleep
Center
Seattle, WA

Sheri Katz, D.D.S.
President of the American Academy of
Dental Sleep Medicine
Darien, IL

Meir Kryger M.D., FRCPC
Physician, Connecticut VA Healthcare
System
Professor, Yale University School of
Medicine
Hartford, CT

Atul Malhotra, M.D.
Associate Professor Medicine, Harvard
Medical School
Medical Director, Sleep Disorders Research
Program
Clinical Chief, Sleep Medicine Pulmonary
& Critical Care and Sleep Medicine
Divisions
Brigham and Women's Hospital
Boston, MA

Tracy R. Nasca
Founder and Senior Vice President
Talk About Sleep, Inc.
Chaska, MN

Allan I. Pack, M.B.Ch.B., Ph.D.
The John Micolot Professor of Medicine
Chief, Division of Sleep Medicine
Director, Center for Sleep and Circadian
Neurobiology
Perelman School of Medicine at the
University of Pennsylvania
Philadelphia, PA

Andrew Scibelli, M.B.A., M.A.
Manager, Employee Health and Well-being
NextEra Energy, Inc.
Juno Beach, FL

Victor G. Villagra, M.D., FACP
Chief Medical Officer
National Rural Electric Cooperative
Association
Arlington, VA

Thomas Weakley
Director of Operations Owner-Operator
Independent Drivers Association (OOIDA)
Foundation, Inc.
Grain Valley, MO

Robert W. Zavoski, M.D., M.P.H.
Medical Director
Medical Care Administration
Connecticut Department of Social Services
Hartford, CT

An anonymous Medical Director.

An anonymous Federal agency stakeholder.

Future Research Needs for Diagnosis of Obstructive Sleep Apnea

Structured Abstract

Background. Obstructive sleep apnea (OSA) is an important public health issue, with challenges for diagnosis and treatment. A recent Comparative Effectiveness Review (CER) found numerous areas with insufficient or low strength of evidence

Purpose. With the assistance of a panel of representative stakeholders, to identify and prioritize future research needs topics for diagnosis of OSA.

Methods. Twenty-one panel members represented six stakeholder categories: patients and the public, providers, purchasers of health care, payers, policymakers, and principal investigators. Building on future research needs topics derived from the CER, stakeholders nominated additional topics for discussion. Nominated topics were discussed by stakeholders (excluding product makers) on a secure Web site discussion board. At the close of the discussion period, stakeholders nominated their top five Future Research Needs topics based on the Agency for Healthcare Research and Quality Effective Health Care Program selection criteria. From these nominations, the highest priority Future Research Needs were determined and were elaborated upon to include possible study designs to address the topics.

Future Research Needs Topics. The high priority future needs topics included:

1. Age and gender specific criteria for abnormal breathing (or OSA)
2. Routine (or selected) preoperative screening for sleep apnea
3. Cost effectiveness of a management strategy (diagnosis [of symptomatic or high-risk patients] through treatment [of patients diagnosed with OSA]), specifically for patients with mild-to-moderate disease severity
 - a. Cost effectiveness of use of diagnostic algorithms and portable monitors, including limited-channel, low-cost portable devices
4. Value of having a sleep medicine specialist involved in the diagnosis of OSA
5. What is the prognostic accuracy of clinical prediction rules to predict clinical outcomes?

Fourteen other future research needs topics were discussed.

Challenges. Stakeholder participation in the online discussion board was low. Discussions were begun by only five stakeholders and only 33 percent of stakeholders participated in the online discussion. The median number of comments across topics was only two. Topic nomination was done by 16 stakeholders (76 percent). Lessons learned from this Future Research Needs panel discussion can be applied to future panels.

Contents

Executive Summary	ES-1
Background	1
Context.....	1
Methods	4
Stakeholder Panel.....	4
Introduction of Process to the Stakeholder Panel	5
Iterative Process To Identify Future Research Needs Topics	6
Use of Microsoft® SharePoint.....	6
Approach to Prioritization.....	9
Approach to Research Question Development and Considerations for Potential Research Designs.....	10
Results	11
Composition of Stakeholder Panel.....	11
Stakeholder Participation and Required Methods Modifications	11
Research Needs	13
High-Priority Future Research Needs Topics	13
High-Priority Future Research Needs Topic 1.....	13
High-Priority Future Research Needs Topic 2.....	16
High-Priority Future Research Needs Topic 3.....	20
High-Priority Future Research Needs Topic 4.....	22
High-Priority Future Research Needs Topic 5.....	25
Second-Tier Future Research Needs Topics.....	28
Second-Tier Future Research Needs Topic 6	28
Second-Tier Future Research Needs Topic 7	28
Other Future Research Needs Topics	28
Other Future Research Needs Topic 8.....	28
Other Future Research Needs Topic 9.....	28
Other Future Research Needs Topic 10.....	29
Other Future Research Needs Topic 11.....	29
Other Future Research Needs Topic 12.....	29
Other Future Research Needs Topic 13.....	29
Other Future Research Needs Topic 14.....	29
Other Future Research Needs Topic 15.....	29
Other Future Research Needs Topic 16.....	29
Other Future Research Needs Topic 17.....	29
Other Future Research Needs Topic 18.....	29
Discussion	30
Challenges in Stakeholder Involvement	30
References	33
Acronyms	34
Tables	
Table A. Evidence Gaps Affecting Conclusions for the Key Questions	ES-2
Table B. List of Future Research Needs Topics	ES-4
Table 1. Evidence Gaps Affecting Conclusions for the Key Questions	3

Table 2. Stakeholders Panel For Diagnosis of Sleep Apnea.....	11
Table 3. List of Future Research Needs Topics	13
Table 4. Sample Size Estimates for Various Clinical Scenarios of Sensitivity and Specificity With an Assumed Prevalence of 10 Percent and an Allowed Variance of 10 Percent in Both Sensitivity and Specificity	27

Figures

Figure A. Analytic Framework for the Diagnosis of Obstructive Sleep Apnea in Adults With Evidence Gaps	ES-1
Figure 1. Analytic Framework for the Diagnosis of Obstructive Sleep Apnea in Adults With Evidence Gaps	2
Figure 2. Sharepoint Homepage: Diagnosis	7
Figure 3. Sample Discussion Thread: Diagnosis	8
Figure 4. Sample Size Calculations For an RCT of Screening Versus no Screening for Postoperative Complications With Varying Estimates of Risk Ratio and Control Rates.....	20

Appendixes

Appendix A. AHRQ’s Effective Health Care Program Selection Criteria for New Research	
--	--

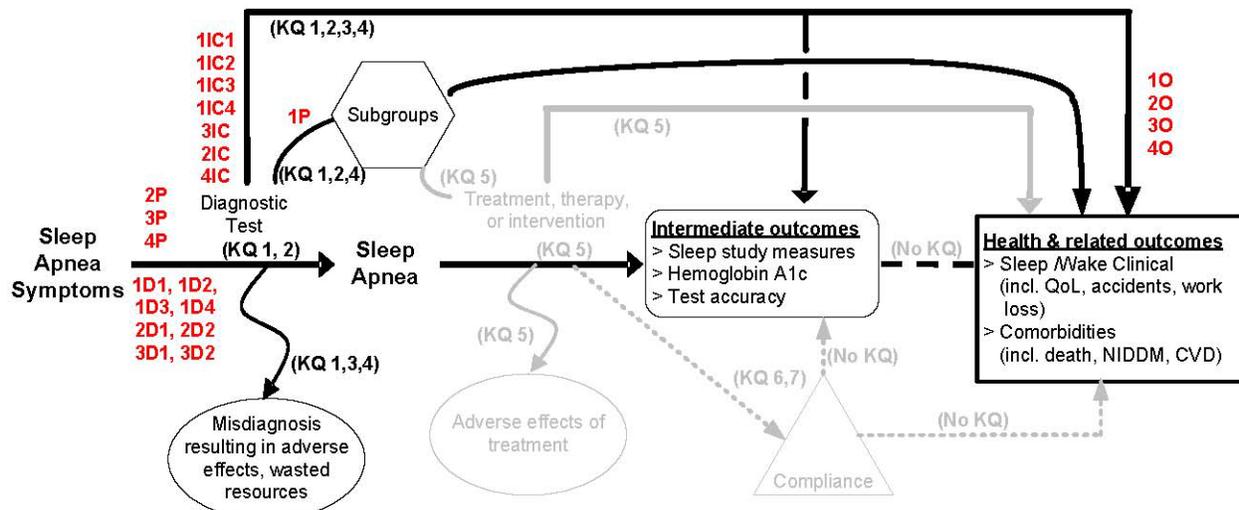
Executive Summary*

Background

The current Future Research Needs (FRN) project was launched shortly after completion of the comparative effectiveness report (CER) on obstructive sleep apnea (OSA). OSA is an important public health issue, due to the considerable mortality and morbidity associated with the condition. The commonly used methods for diagnosing and treating OSA are cumbersome, resource-intensive, and often inconvenient for the patient. The Tufts EPC conducted a CER on diagnostic tools, characteristics of OSA that are predictive of poor outcomes, and treatments for OSA. For the purpose of the FRN process, the original OSA CER was divided into two overarching sections: diagnosis and treatment. This document describes the FRN for diagnosis of OSA; an accompanying parallel report describes the FRN for treatment. For the most part, the Background, Methods, and description of the challenges are nearly identical between the two reports.

Figure A is an analytic framework to visualize areas of the systematic review in which evidence gaps were identified. Table A summarizes the evidence gaps identified in our review of the diagnosis of OSA in adults (the CER's Key Questions 1–4). These gaps in the evidence review limited the ability to make conclusions on the questions asked; thus they formed the initial FRN topics.

Figure A. Analytic framework for the diagnosis of obstructive sleep apnea in adults with evidence gaps



CVD = cardiovascular disease; D = study design; IC = intervention and comparator; KQ = Key Question; NIDDM = noninsulin dependent diabetes mellitus; O = outcome; P = population; QoL = quality of life

The analytic framework above highlights the evidence gaps in red that were identified as affecting the conclusions for the respective Key Questions in the CER. The alphanumeric code for the gaps corresponds to the detailed gaps that are listed in Table 1. The first number of the code corresponds to the key question, the following letters represent the PICOD domains, and the last numerical corresponds to the number on the list for that particular Key Question and domain. Where there is only one gap identified, the last number is dropped. Grayed out portions of the analytic framework are treatment-related questions that are covered in the companion report.³

* Please refer to the main report for references.

Table A. Evidence gaps affecting conclusions for the Key Questions

Key Question	Category	Evidence Gap
Key Question 1. How do different available tests compare in their ability to diagnose sleep apnea in adults with symptoms suggestive of disordered sleep? Key Question 1a. How do these tests compare in different subgroups of patients, based on: race, sex, body mass index, existing noninsulin dependent diabetes mellitus, existing cardiovascular disease, existing hypertension, clinical symptoms, previous stroke, or airway characteristics?	Population	1P: No subgroup analyses available for any test - none of the studies explicitly evaluated the monitors in patients with important comorbidities such as chronic lung disease, or congestive heart failure.
	Intervention / Comparator	1IC1: No head-to-head comparisons of portable monitors, questionnaires, and prediction rules. 1IC2: Insufficient data to make conclusions about the following questionnaires: STOP, STOP-Bang, ASA checklist, Hawaii Sleep Questionnaire. 1IC3: No tests based on severity of symptoms. 1IC4: No standardized method of testing used.
	Outcomes	1O: No analyses of clinical outcomes, including response to treatment or process outcomes.
	Design	1D1: Incomplete reporting and inadequate analyses were common. 1D2: Few high quality studies with little susceptibility to bias. 1D3: The studies did not allow us to adequately assess any issues related to night-to-night variation. 1D4: Verification bias (i.e. not testing all participants with all the devices that are being compared) is seen in many studies. 1D5: Most of the studies performed at academic/research centers. It is not clear how the results would generalize to the general population.
Key Question 2. How does phased testing (screening tests or battery followed by full test) compare to full testing alone?	Population	2P: Studies on a general population of people with OSA are needed.
	Intervention / Comparator	2IC: No studies directly address this question.
	Outcomes	2O: No analyses of clinical outcomes were done.
	Design	2D1: One available study subject to verification bias. 2D2: Studies evaluating the appropriateness of tests based on patient characteristics and severity of their symptoms were not available.
Key Question 3. What is the effect of preoperative screening for sleep apnea on surgical outcomes?	Population	3P: More studies on general surgical patients are needed.
	Intervention / Comparator	3IC: A broad range of screening tools, including questionnaires; need to be evaluated against PSG.
	Outcomes	3O: Long-term outcomes were not evaluated.
	Design	3D1: No randomized trials address this question. 3D2: Selection biases of major concern in available studies.
Key Question 4. In adults being screened for obstructive sleep apnea, what are the relationships between apnea-hypopnea index or oxygen desaturation index, and other patient characteristics with long-term clinical and functional outcomes?	Population	4P: Further studies evaluating the link between apnea-hypopnea index, and diabetes mellitus and hypertension are needed.
	Intervention / Comparator	4IC: Only AHI was well-studied, not other indices and patient characteristics.
	Outcomes	4O: Other than all-cause mortality, clinical outcomes are understudied.

AHI = apnea-hypopnea index; ASA = American Society of Anesthesiologists; OSA = obstructive sleep apnea; PSG = polysomnography; STOP = Snoring; STOP-Bang = STOP with body mass index, age, neck circumference, and sex variables

Methods

Stakeholder Involvement

FRN topics nominated for prioritization were identified based on recommendations from the original CER and through an iterative process utilizing a diverse stakeholder panel. After a formal recruitment process, participating panelists were asked to review, discuss and nominate for prioritization CER-derived FRN topics, as well as submit additional topics for FRN consideration. The panel consisted of patient advocates, providers, purchasers, payers, policymakers, researchers and product makers, representing the full range of stakeholders who may use research evidence in health care and public health decisionmaking.

Use of Microsoft® SharePoint

The Evidence-based Practice Center (EPC) developed a secure, password-protected Microsoft® SharePoint Web site primarily to host the online stakeholder discussions of the FRN topics. The SharePoint Web site also functioned as a platform for Tufts EPC staff to post project announcements and reference documents.

Throughout the open discussion period, stakeholders could use SharePoint to submit new FRN topics; a direct email to EPC staff was also an option. All stakeholder-submitted topics were reviewed and refined by the EPC. These topics, along with the CER-derived FRN topics, were posted by EPC staff on the SharePoint Web site for stakeholder discussion.

Each FRN topic was posted as a separate “discussion board.” For their discussions, stakeholders were asked to consider four dimensions of need related to the proposed topic: (1) importance; (2) desirability of research/duplication; (3) feasibility; and (4) potential impact. Stakeholder participation was encouraged and monitored throughout the discussion period, and discussion boards were moderated daily by Tufts EPC staff to ensure appropriateness and relevance of all comments.

Approach to Prioritization

After the close of the online discussion period, stakeholders were asked to identify up to five FRN topics—considering the topic’s importance, desirability, feasibility and potential impact—that were of highest priority and that met the Effective Health Care Program Selection Criteria. Nomination was conducted by email and individual phone calls. After nomination, the EPC grouped similar topics into overarching topics and categorized them into four groups: (1) high-priority topics; (2) second-tier priority topics; (3) topics of little interest to stakeholders; and (4) topics not meeting Effective Health Care Program appropriateness criteria.

Research Question Development and Considerations for Potential Research Design

For each “high-priority” FRN topic, EPC staff considered the range of study designs that would best address the topic. To determine candidate study designs, the feasibility of the study designs, and sample size calculations, we followed the structure laid out in the Future Research Needs document Framework for Considering Study Designs for Future Research Needs.⁴ For each topic, we described our assumptions about the most appropriate PICOD criteria (Population, Intervention, Comparator, Outcomes, study Design), in particular describing the advantages and disadvantages of various potential research designs. We specifically considered

the feasibility of the research questions focusing on potential sample size, time, and recruitment issues.

Results

Table B lists the final FRN topics.

Table B. List of future research needs topics

High-Priority Future Research Needs Topics
1. Age- and gender-specific criteria for defining the OSA syndrome (and abnormal breathing)
2. Effect of routine (or selected) preoperative screening for sleep apnea
3. Cost-effectiveness analysis of a management strategy (diagnosis of symptomatic or high-risk patients through treatments of patients diagnosed with OSA), specifically for patients with mild to moderate disease severity
4. Value of having a sleep medicine specialist involved in the diagnosis of OSA (in addition to or instead of a nonspecialist)
5. What is the prognostic accuracy of clinical prediction rules (CPRs) to predict clinical outcomes?
Second-Tier Future Research Needs Topics
6. Indications (patient signs, symptoms, or other features) for appropriate home testing
7. Diagnostic approaches to OSA in obese and nonobese patients
Other Future Research Needs Topic
8. Can PSG be skipped in making the diagnosis of sleep apnea?
9. What are the financial barriers to access to diagnosis?
10. Head-to-head comparisons of portable monitors, questionnaires, and prediction rules
11. Association between use of questionnaires and clinical outcomes
12. What are the available, objectively-measured predictors of sleep apnea diagnosis?
13. What is consumer willingness-to-pay for screening, to identify consumer preferences for strategies to diagnose sleep apnea?
14. Value of scoring nasal flow limitation in recognizing mild OSA
15. Value of brain MRI in evaluating OSA patients
16. Randomized trials of phased testing
17. Value of using 4-phase rhinomanometry in recognition of patients with high nasal resistance and OSA
18. Diagnostic approach to OSA in micrognathia and retrognathia

High-Priority Future Research Needs Topic 1

Age- and gender-specific criteria for defining the OSA syndrome (and abnormal breathing)

This overarching topic was addressed in general terms in the Introduction of the CER, but it was not a Key Question in the report. A Technology Assessment preceding the CER concluded that most experts consider laboratory-based PSG as the reference method to identify people with apnea-hypopnea index (AHI) suggestive of OSA. However, this does not mean that facility-based PSG is an error-free “gold standard” for measuring (abnormal) breathing, or that facility-based PSG measurements of breathing are generally sufficient for defining the OSA syndrome. The stakeholders stated that there is a need for a new definition of the OSA syndrome that identifies individuals with breathing abnormalities during sleep who have or are at risk of developing clinical sequelae because of their exposure to sleep-disordered breathing. The aim of this future research need topic is to identify age- and gender-specific criteria for defining individuals who are at increased health risk because of abnormal breathing during sleep (OSA syndrome); it is not to discuss the clinical utility of (yet undefined) age- and gender-specific diagnostic criteria.

Prospective Cohort Study

Prospective longitudinal cohort studies are the most informative studies to assess predictors of natural history outcomes (e.g., age and sex as criteria for OSA). A well-designed prospective study will be less biased than a retrospective database analysis. Randomized trials of interventions generally do not provide better data to assess predictors of natural history outcomes. We suggest a prospective cohort study of people selected from the general population to associate measurements of breathing with a battery of short-term pathophysiological measurements that distinguish people whose breathing patterns have immediate physiological impact, from those whose breathing patterns do not have measurable functional sequelae. Multivariable analyses that control for potential confounders, such as comorbidity and body mass index, should be performed to evaluate the relationship between measurements of breathing (e.g., AHI scores) and functional outcomes by sex and age groups. It should be noted that the definitive definition of OSA will remain unclear because whether the functional outcomes are good proxy markers for clinical outcomes will likely remain unknown. With long-term followup, the proposed cohort study can also provide natural history data. Any subsequent treatments and additional diagnostic testing should also be recorded as these data could inform the clinical utility of the age- and gender-specific diagnostic criteria.

High-Priority Future Research Need Topic 2

Effect of routine (or selected) preoperative screening for sleep apnea

The CER found insufficient evidence to address this topic. The stakeholders' discussions centered on the anecdotal nature of the evidence regarding perioperative risks in patients with undiagnosed OSA. Other issues that were identified as lacking evidence related to patient risk characteristics, the type of surgery, and the outcomes of interest. There is a need to evaluate the value of routine (or selected) preoperative screening for OSA, to assess whether it is justified as a part of routine preoperative assessment.

Randomized Controlled Trials (RCTs)

A randomized comparison of screening for OSA with extended anesthesia care when appropriate versus no screening with routine anesthesia care would provide information for effectiveness of the screening protocol. The design of choice would likely be a multicenter cluster-randomized trial, where whole centers would be randomized to screen patients or not. The advantage of using a cluster-randomized trial is that there is a minimal risk of protocol deviation if active OSA screening is performed as part of a center's clinical protocols. There would also be little risk of cross-contamination, of OSA screening to patients who had been randomized to not receive screening. It is also likely that recruitment and randomization will be logistically easier if it is done at the center level, rather than at the patient level within a center. Use of validated screening questionnaires and the shorter perioperative follow up period make it easier to recruit. However, since the event rate for postoperative complications is low due to the advances in perioperative care, the required sample sizes are number in the thousands, ranging from 2000 to 100,000. These could result in the use of large resources in terms of cost, time and effort, and the decision to use these resources for conducting trials has to be balanced against the benefit of screening on postoperative complications. Loss to followup should be rare if an adequate system were in place to follow patients. If a convincing argument can be made that a continuous outcome could be an adequate proxy outcome for postoperative complications, then it

is likely that a smaller sample size would be needed for adequate power for this outcome. However, it is currently unclear than any continuous outcome would be a convincing proxy.

Observational Studies

Observational studies, such as those that exist, are likely to be fundamentally flawed. In normal clinical practice, preoperative patients who undergo testing for OSA will always be greatly different than unscreened patients. However, these studies would be less resource-intensive than a trial. The validity of the data collected using a flawed study design detracts from the benefit gained by decreased resource utilization.

High-Priority Future Research Needs Topic 3

Cost-effectiveness analysis of a management strategy (diagnosis of symptomatic or high-risk patients through treatments of patients diagnosed with OSA), specifically for patients with mild to moderate disease severity

Cost-effectiveness analysis allows the comparison of different interventions on similar benefit, cost and utility scales. Benefit and cost estimates can have both internal and external validity. The objective of this FRN is to establish better evidence about the costs and benefits of alternative diagnostic strategies for individual patients with mild to moderate disease severity. The companion report presents a similar discussion on the costs and benefits of alternative treatment strategies.

Systematic Review

Cost-effectiveness analyses were not addressed by the CER. Conducting a systematic review may be the first step to ascertain the level of existing evidence.

Cost–Benefit Analysis

A quality-adjusted cost–benefit analysis is recommended, comparing the incremental costs and benefits of different diagnostic strategies to each other. Benefit, utility, and cost estimates may be derived from previous clinical studies data, where available, and otherwise from observational data. These estimates should include clinical, work-related, accident, and quality of life outcomes. Out-of-pocket patient costs should also be included in cost estimates. A societal perspective should be assumed in the main analysis and the patient perspective in a sub-analysis. A variety of diagnostic strategies should be considered, including but not limited to PSG in everyone, diagnostic algorithms and clinical prediction rules, portable monitoring, and phased testing. Other tests could include four-phase rhinomanometry and brain magnetic resonance imaging. Because a cost-effectiveness analysis can draw from previously collected data, the cost, size and duration of such studies can be limited. However, as discussed by the stakeholders, it may be challenging to gather all the relevant data.

High-Priority Future Research Needs Topic 4

Value of having a sleep medicine specialist involved in the diagnosis of OSA

Increasingly patients with OSA are being diagnosed by primary care providers. Home studies have allowed the primary care physician to bypass the sleep center altogether, and national companies have started marketing the use of home studies to primary care providers.

Systematic Review

The CER did not address the effect of having different specialists involved in care, thus a systematic review may be the first step to ascertain the level of existing evidence.

Analyzing Claims Data

Analyzing claims data provided by an insurance provider or a health care system, such as from deidentified Medicare and/or Medicaid claims data, could help to ascertain cost differences with use of a specialist. While this approach would be able to provide a large sample of patients from diverse geographic locations, it would not be a controlled analysis.

Post Hoc Analysis of Existing Trials

Another study option would be to review existing studies and analyze any available information on diagnosis by a sleep specialist versus diagnosis by a nonspecialist. This approach may require additional unpublished data from study authors.

Survey of Providers

A cross-sectional survey of providers would provide an in-depth view of issues associated with using a sleep specialist in the diagnosis of OSA. However, the size of the survey would be limited by participation rates, and the cross-sectional design would not be able to answer issues of causation.

High-Priority Future Research Needs Topic 5

What is the prognostic accuracy of clinical prediction rules to predict clinical outcomes?

The CER found a low strength of evidence among seven studies that some clinical prediction rules (CPRs) may be useful in the prediction of a diagnosis of OSA. However, none of the studies examined this topic, namely whether use of a CPR resulted in improved clinical outcomes. The aim of the topic is to evaluate the “prognostic” accuracy of existing CPRs, to evaluate, whether CPRs can determine who will experience a clinical outcome in the future. Different CPR thresholds will correspond to different counts of patients falling into different risk (or prognostic) groups based on clinical outcomes; the number of clinical events in the risk categories identified by the thresholds will vary depending on the actual value of the threshold used. Arguably, the most appropriate study design to address the prognostic accuracy of CPRs is a prospective observational study.

Prospective Observational Study

Prospective observational studies are best suited to assess the prognostic value of a CPR because they can study multiple risk factors of interest as well as clinical outcomes in a general patient population in whom the CPR will eventually be used, ensuring applicability and external validity of the results. The CPR would be applied to participants upon their entry to the cohort, and they would be followed to ascertain whether they experience an outcome or not. Another design is a nested case-control study, where all patients in a defined cohort who experience an event are designated as “cases” and are matched to a collection of “controls” (i.e., cohort participants who did not experience the outcome). A major advantage of a nested case-control study is that it allows an adequately powered post hoc analysis of prospectively collected data where the event rate is relatively rare. If suitable databases are immediately available, one can perform a case-control study in a relatively short time, as there is no need to wait for followup. However, it may be difficult to reconstruct the CPR or identify people who match the setting of interest from a retrospective database, and therefore prospective collection of data may be unavoidable. Prospective cohorts would take more time and effort compared to a case-control study built on an existing database. For example, when assessing incident cardiovascular disease or diabetes mellitus, the timeframe could be months or years. When assessing mortality outcomes, the timeframe could very well extend to decades. Because of the substantial resources necessary for a prospective cohort study or a nested case-control study, it is probably not practical to design a study whose sole purpose is to assess the prognostic ability of CPRs. Instead, it would be preferable to incorporate the assessment of the prognostic performance of CPRs into a prospective cohort study in which assessment of prognostic performance will be one of several aims. For different scenarios of using CPRs with differential sensitivity and specificity pairs, the number of subjects that need to be studied ranges in the hundreds for nested case control studies to a few thousands for prospective cohorts. Patient recruitment should be straightforward and relatively simple, as patients are interested in knowing whether they have a condition that is known to cause complications and is associated with chronic disease outcomes.

Discussion

We implemented a Web-based discussion board in preference over a series of teleconferences because of what we believed would be advantages of the online approach. These include: greater flexibility for stakeholders; full participation by all stakeholders in all discussions; a full record of all discussions; less time expenditure by stakeholders; and less resource expenditure by EPC staff. However, we encountered low participation rate of stakeholders during most stages of the project. In contrast, conducting a 1.5-hour teleconference for three consumer stakeholders produced six topic nominations, highlighting the difference in impact between the two communication platforms. Overall, although SharePoint offered the convenience of asynchronous collaboration it seemed to lack appropriate incentives to engage stakeholders.

Several possible solutions exist for these problems, including reducing the number of stakeholders; reverting to conducting a series of teleconferences; using focus groups to allow a full, simultaneous discussion of each topic; combining teleconferences with online discussions; or numerous other similar approaches. Regardless of which type of discussion is held, there may be advantages to limiting the size of the stakeholder panel and to choose stakeholders who show enthusiasm in joining the panel and participating in discussions.

The stakeholder discussion identifying FRN topics on diagnosis of OSA highlights the vast field of research that could potentially be done in this area. Three out of the five nominated topics fell outside the purview of the questions addressed in the CER. The scope of the original CER was focused on the use of portable tools and screening, phased testing and preoperative screening, and the relationships of OSA indices to clinical outcomes. Notably, the stakeholders considered a range of topics broader than the CER's Key Questions regarding diagnosis of OSA to be important for future research.

In summary, the online Web site discussion of numerous topics by a large stakeholder panel was only moderately successful. Various approaches to improve participation and discussion are possible, including increased use of teleconferences, restricting the size and members of the stakeholder panel, and other approaches. More experience with different approaches is needed.

Background

Context

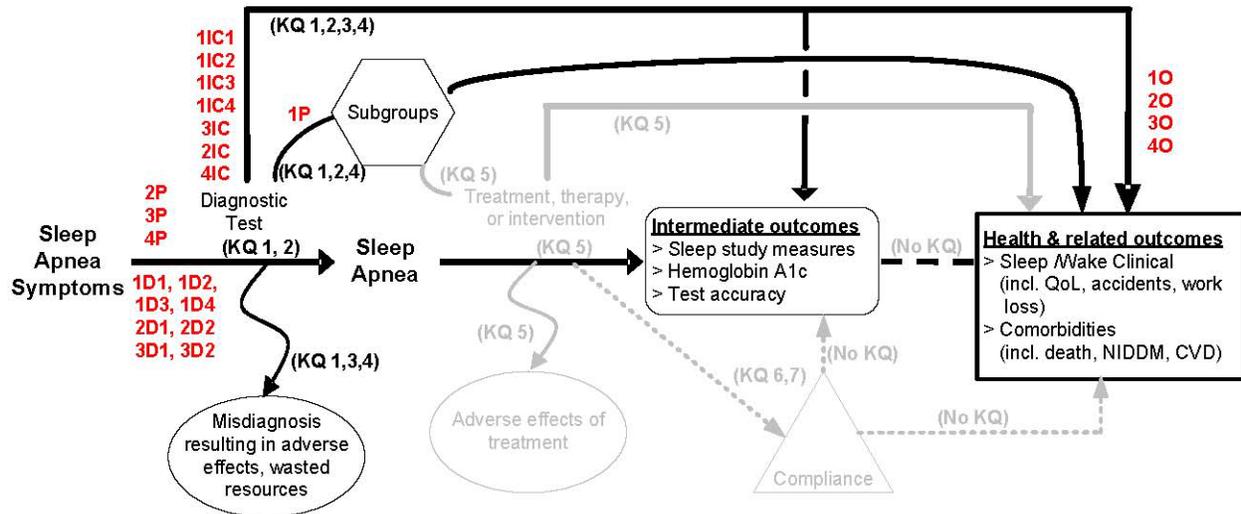
Comparative Effectiveness Reviews (CER)—systematic reviews of existing research on the effectiveness, comparative effectiveness, and comparative harms of different health care interventions—are intended to provide relevant evidence to inform real-world health care decisions for patients, providers, and policymakers. In addition to synthesizing the evidence, CERs also identify the gaps in evidence that limit the ability to answer the key research questions. As part of an effort beginning in 2010, the Agency for Healthcare Research and Quality (AHRQ) supports its Evidence-based Practice Centers (EPCs) to work with various stakeholders to further develop and prioritize the future research needed by decisionmakers.² This process is new, and the methods to delineate future research needs (FRN) are not yet fully developed. The current report describes the first experience of the Tufts EPC in a stakeholder-driven process to identify and nominate for prioritization FRN topics (other than a pilot process with a limited involvement of a small number of stakeholders).

The FRN document is intended to inform and support researchers and those who fund research to ultimately enhance the body of comparative effectiveness evidence so that it is useful for decisionmakers. This document describes the process of developing a prioritized list of research needs with considerations of the advantages and disadvantages of various potential research designs to help researchers and funders develop future research proposals or solicitations, respectively. This process begins with identification of evidence gaps from the original CER, followed by the addition of other areas potentially requiring further research, nomination for prioritization of these evidence gaps by stakeholders, and development of potential study designs for the highest priority topics. Although researchers and funders of research are the end-users of the report, the resulting research is meant to improve health care decisions; therefore, the stakeholders for this process include patients, clinicians, research investigators, payers, and policymakers.

The current FRN project was launched shortly after completion of the CER on obstructive sleep apnea (OSA). OSA is an important public health issue, due to the considerable mortality and morbidity associated with the condition. The commonly used methods for diagnosing and treating OSA are cumbersome, resource intensive, and often inconvenient for the patient. The Tufts EPC conducted a CER on diagnostic tools, characteristics of OSA that are predictive of poor outcomes, and treatments for OSA.¹ For the purpose of the FRN process, the original OSA CER was divided into two overarching sections: diagnosis and treatment. This document describes the FRN for diagnosis of OSA; an accompanying parallel report describes the FRN for treatment.³ For the most part, the Background, Methods, and description of the challenges are nearly identical between the two reports.

Figure 1 is an analytic framework to visualize areas of the systematic review in which evidence gaps were identified. Table 1 summarizes the evidence gaps identified in our review of the diagnosis of OSA in adults (the CER's Key Questions 1–4). These gaps in the evidence review limited the ability to make conclusions on the questions asked; thus they formed the initial FRN topics.

Figure 1. Analytic framework for the diagnosis of obstructive sleep apnea in adults with evidence gaps



CVD = cardiovascular disease; D = study design; IC = intervention and comparator; KQ = Key Question; NIDDM = noninsulin dependent diabetes mellitus; O = outcome; P = population; QoL = quality of life;

The analytic framework above highlights the evidence gaps in red that were identified as affecting the conclusions for the respective Key Questions in the CER. The alphanumeric code for the gaps corresponds to the detailed gaps that are listed in Table 1. The first number of the code corresponds to the Key Question, the following letters represent the PICOD domains, and the last numerical corresponds to the number on the list for that particular Key Question and domain. Where there is only one gap identified, the last number is dropped. Grayed out portions of the analytic framework are treatment-related questions that are covered in the companion report.³

Table 1. Evidence gaps affecting conclusions for the Key Questions

Key Question	Category	Evidence Gap
Key Question 1. How do different available tests compare in their ability to diagnose sleep apnea in adults with symptoms suggestive of disordered sleep? Key Question 1a. How do these tests compare in different subgroups of patients, based on: race, sex, body mass index, existing noninsulin dependent diabetes mellitus, existing cardiovascular disease, existing hypertension, clinical symptoms, previous stroke, or airway characteristics?	Population	1P: No subgroup analyses available for any test—none of the studies explicitly evaluated the monitors in patients with important comorbidities such as chronic lung disease, or congestive heart failure.
	Intervention / Comparator	1IC1: No head-to-head comparisons of portable monitors, questionnaires, and prediction rules. 1IC2: Insufficient data to make conclusions about the following questionnaires: STOP, STOP-Bang, ASA checklist, Hawaii Sleep Questionnaire. 1IC3: No tests based on severity of symptoms. 1IC4: No standardized method of testing used.
	Outcomes	1O: No analyses of clinical outcomes, including response to treatment or process outcomes.
	Design	1D1: Incomplete reporting and inadequate analyses were common. 1D2: Few high quality studies with little susceptibility to bias. 1D3: The studies did not allow us to adequately assess any issues related to night-to-night variation. 1D4: Verification bias (i.e. not testing all participants with all the devices that are being compared) is seen in many studies. 1D5: Most of the studies performed at academic/research centers. It is not clear how the results would generalize to the general population.
Key Question 2. How does phased testing (screening tests or battery followed by full test) compare to full testing alone?	Population	2P: Studies on a general population of people with OSA are needed.
	Intervention / Comparator	2IC: No studies directly address this question.
	Outcomes	2O: No analyses of clinical outcomes were done.
	Design	2D1: One available study subject to verification bias. 2D2: Studies evaluating the appropriateness of tests based on patient characteristics and severity of their symptoms were not available.
Key Question 3. What is the effect of preoperative screening for sleep apnea on surgical outcomes?	Population	3P: More studies on general surgical patients are needed.
	Intervention / Comparator	3IC: A broad range of screening tools, including questionnaires; need to be evaluated against PSG.
	Outcomes	3O: Long-term outcomes were not evaluated.
	Design	3D1: No randomized trials address this question. 3D2: Selection biases of major concern in available studies.
Key Question 4. In adults being screened for obstructive sleep apnea, what are the relationships between apnea-hypopnea index or oxygen desaturation index, and other patient characteristics with long-term clinical and functional outcomes?	Population	4P: Further studies evaluating the link between apnea-hypopnea index, and diabetes mellitus and hypertension are needed.
	Intervention / Comparator	4IC: Only AHI was well studied, not other indices and patient characteristics.
	Outcomes	4O: Other than all-cause mortality, clinical outcomes are understudied.

AHI = apnea-hypopnea index; ASA = American Society of Anesthesiologists; OSA = obstructive sleep apnea; PSG = polysomnography; STOP = snoring; STOP-Bang = STOP with body mass index, age, neck circumference, and sex variables.

Methods

The Methods section describes the a priori protocol methods as planned prior to convening the stakeholder panel. Much of the process did not go as planned so several major modifications were made to the protocol during the process of developing the Future Research Needs (FRN) document. These modifications are described in the first part of the Results section. A description of lessons learned and suggestions for further modifications to the methods are in the Discussion section.

We used an iterative process with a stakeholder panel to identify FRN topics for prioritization. From the original Comparative Effectiveness Review (CER), the Evidence-based Practice Center (EPC) generated an initial list of FRN topics and then solicited additional topics from the stakeholder panel.

Stakeholder Panel

We considered seven stakeholder categories to build a panel representing the full range of stakeholders who may use research evidence in health care and public health decisionmaking.

1. Patients and the public: This group represents current and potential consumers of patient-centered health care and population-focused public health. This group also includes caregivers, family members, and patient advocacy organizations, all of whom address the interests of consumers.
2. Providers: This group includes individuals (e.g., nurses, physicians, mental health counselors and other providers of care and support services) and organizations (e.g., hospitals, clinics, community health centers, community-based organizations, pharmacies, EMS agencies, skilled nursing facilities, schools) that provide care to patients and populations.
3. Purchasers: This group includes employers, the self-insured, government, and other entities responsible for underwriting the costs of health care.
4. Payers: This group represents insurers, Medicare and Medicaid, individuals with deductibles, and others responsible for reimbursement for interventions and episodes of care.
5. Policymakers: This group includes organizations such as the White House, the U.S. Department of Health and Human Services, Congress, States, professional associations, and intermediary groups that collate and distribute information to policymakers.
6. Principal investigators, researchers, and research funders.
7. Product makers: This group represents drug and device manufacturers.

These categories are not necessarily mutually exclusive. In concept, each individual represents at least one key type of health care and health care research decisionmaker. Any single person or entity may have several roles and may be responsible for different types of decisions. For example, some health care purchasers are also payers, and conversely, some payers also provide care. Patients and their advocates may be providers or employers with policymaking responsibilities, and so on. In addition, each of these seven stakeholder types may be focused on applying CER at the patient level or at the population level. Patient-level decisions include questions pertaining to what treatment would be best for a given patient at a particular time.

Population-level decisions include questions pertaining to what services, resources, policies or other alternatives are best for groups of patients and entire communities that are connected by practice setting, geography, clinical domain or other cluster. To be patient centered, decisions made about groups of patients must recognize both the diversity of needs across populations and the heterogeneity of individuals within populations.

Identification and Invitation of Individual Stakeholders

We compiled a list of potential stakeholders from three sources: (1) those individuals who were or had been invited to be a member of the sleep apnea report's Key Informant panel or Technical Expert Panel, or a peer reviewer; (2) a previously compiled list of stakeholders assembled for a 2010 Stakeholder Forum on Comparative Effectiveness Research for the National Institutes of Health Clinical and Translational Science Awards program; and (3) a list of stakeholders compiled in 2010 for the AHRQ Effective Health Care Program. We also solicited recommendations for stakeholders from selected government agencies, professional organizations, and other representative bodies. We selected people who would potentially fall within one or more of the stakeholder categories for either obstructive sleep apnea (OSA) diagnosis or treatment. With the assistance of our local domain expert, we initially selected the most promising individuals based on their perceived interest in the topic, their level of previous participation in discussions on the topic, and their fit into the stakeholder categories.

Individuals who met criteria to be a stakeholder were contacted directly by email with a brief description of the project, an invitation to be a stakeholder for one or both stakeholder panels (diagnosis and treatment), and the Executive Summary of our sleep apnea CER. Potential stakeholders were also telephoned, as necessary, to solicit their interest. Other individuals on our list were contacted specifically requesting suggestions for appropriate stakeholders. These people, in turn, were contacted by email. In addition, through Internet searches and focused searches in MEDLINE, we found other potential stakeholders or individuals who could suggest stakeholders. In particular, we contacted senior level administrators at various governmental, nongovernmental, and professional organizations, including but not limited to, the National Association of Community Health Centers, the Veterans Administration, Blue Cross Blue Shield, Centers for Medicare and Medicaid Services, Academy Health, the National Heart, Lung, and Blood Institute, the Office of Minority Health, and the Centers for Disease Control and Prevention. Patients were solicited from our Key Informants list, personal contacts, and an Internet search of sleep apnea advocacy, support, or discussion groups. Selected potential stakeholders were invited to participate in both Future Research Needs Prioritization projects on OSA (diagnosis and treatment), which were run concurrently. All stakeholders completed a standard disclosure of interest form.

We also contacted manufacturers from the list of companies that were sent Scientific Information Packets for the original sleep apnea report. These companies were asked to provide potential FRN topics with rationales, but were not invited to be stakeholders.

Introduction of Process to the Stakeholder Panel

Along with an invitation letter, we distributed the executive summary and the Future Research section of the original CER¹ to the invited stakeholders. The original Key Questions, summary of evidence table, and the Implications for Future Research sections in the executive summary were highlighted. The purpose of the FRN project and expectations for the input from the stakeholders were outlined clearly in the invitation letter.

We scheduled teleconferences to allow general introductions and for the EPC to explain the purpose and process of the FRN topic development process, after compiling the stakeholder panel.

Iterative Process To Identify Future Research Needs Topics

The EPC reviewed the CER Future Research section and, from this, developed a series of FRN topics. We wrote a brief rationale statement for each. These formed the initial list of FRN topics.

After stakeholders submitted their disclosure of interest statements, we invited them to submit FRN topics to the EPC. For each FRN topic, stakeholders were asked to provide a brief rationale (maximum 250 words) considering the four dimensions of need as listed below under Approach to Prioritization. We reviewed submitted FRN topics, and planned to categorize each topic into one of three categories: “definitely relevant to Key Questions,” “not relevant to Key Questions,” or “unclear.” We planned to combine duplicate or similar FRN topics together into one topic. For FRN topics categorized as “unclear,” we asked the stakeholders to provide additional information or clarifications. All topics were distributed to stakeholders. Throughout the stakeholder panel discussion period, stakeholders were invited to submit new FRN topics.

Use of Microsoft® SharePoint

Two separate Microsoft SharePoint Web sites were created to: (1) host the FRN discussion; (2) submit additional topics for discussion; and (3) nominate topics for future research. SharePoint resources were provided through the Tufts Clinical and Translational Science Institute, of which the Tufts EPC is a member. One Web site was created for the diagnosis topics discussed in this document and a separate Web site for the parallel treatment discussion. SharePoint was chosen as it offered stakeholders the flexibility of an asynchronous online discussion forum. Additionally, the secure Web site provided Tufts staff the most control, in terms of site content, structure and functionality, when compared with other publicly available discussion platforms. The secure, password-protected Web site was housed behind the Tufts firewall and accessible to stakeholders via invitation only. The SharePoint Web site also served as a platform for Tufts EPC staff to post project announcements, a welcome video, and important reference documents. These documents included instructions to stakeholders on how to navigate and use the Web site, an FRN project overview, the Executive Summary and Key Questions of the sleep apnea CER, lists of participating stakeholders, initial FRN discussion topics, common abbreviations and acronyms, announcements, and reference documents. The documents on the Web site were also emailed to all stakeholders. In addition to the EPC staff and the stakeholders, the AHRQ Task Order Officer was given access to the Web site.

The FRN topic discussion boards were the primary feature of the SharePoint Web site. Discussion boards were prominently placed front and center on the Web site’s homepage (Figure 2). Links to individual discussion boards were also strategically placed on the Web site’s main navigation toolbar for direct access from any page on the Web site.

Figure 2. SharePoint homepage: Diagnosis



Sleep Apnea Future Research Needs: Diagnosis

All Sites

View All Site Content

Home Page

Sleep Apnea Treatment Website

Contact Tufts Staff

Submit FRN Topic for Discussion

Sleep Apnea Diagnosis Topic Discussions

- 1. Association between use of clinical prediction rules and clinical outcomes
- 2. Association between use of questionnaires and clinical outcomes
- 3. Head-to-head comparisons of portable monitors, questionnaires, and prediction rules
- 4. Randomized trials of phased testing
- 5. Routine (or selected) preoperative screening for sleep apnea
- 6. Defining the OSA syndrome
- 7. Value of having a sleep medicine specialist involved in the diagnosis of OSA
- 8. Indications (patient signs, symptoms, or other features) for appropriate home testing
- 9. Value of brain MRI in evaluating OSA patients
- 10. Value of scoring nasal flow limitation in recognizing mild OSA
- 11. Age and gender

Sleep Apnea Diagnosis FRN Discussion Board (click on topic below to join discussion)

Subject	Replies
2. Association between use of questionnaires and clinical outcomes	4
1. Association between use of clinical prediction rules and clinical outcomes	9
3. Head-to-head comparisons of portable monitors, questionnaires, and prediction rules	4
4. Randomized trials of phased testing	3
5. Routine (or selected) preoperative screening for sleep apnea	10
6. Defining the OSA syndrome	4
7. Value of having a sleep medicine specialist involved in the diagnosis of OSA	5
8. Indications (patient signs, symptoms, or other features) for appropriate home testing	2
9. Value of brain MRI in evaluating OSA patients	4
10. Value of scoring nasal flow limitation in recognizing mild OSA	3
11. Age and gender specific criteria for abnormal breathing (or OSA)	3
12. Diagnostic approaches to OSA in obese and non-obese patients	3
13. Value of using 4-phase Rhinomanometry in recognition of patients with high nasal resistance and OSA	2
14. Cost-effectiveness of a management strategy (diagnosis [of symptomatic or high-risk patients] through treatment [of patients diagnosed with OSA]), specifically for patients with mild-to-moderate disease severity	3
15. Cost effectiveness of use of diagnostic algorithms and portable monitors, including limited-channel, low cost portable devices	1
16. Can PSG be skipped in making the diagnosis of sleep apnea?	4
17. What are the financial barriers to access to diagnosis?	0
18. What is consumer willingness-to-pay for screening, to identify consumer preferences for strategies to diagnose sleep apnea?	0
19. What are the available, objectively-measured predictors of sleep apnea diagnosis?	1

Site Actions

Announcements

4. Sleep Apnea report now available at AHRQ website 8/10/2011 1:10 PM
by tuftsepc
The complete report on the diagnosis and treatment of OSA is now available on the AHRQ website. It is available in various forms - the full report, the executive summary, the Clinicians Guide, and the Consumers Guide. Click the link titled "AHRQ..."

3. Presentation about FRN project added to Reference Documents 8/5/2011 4:37 PM
by tuftsepc
Thank you for your participation in the conference calls. For participants who were unable to join the two conference calls we had this week, we have uploaded the presentation made at the call, titled "Overview of Future Research needs project" in...

1. Welcome to the Sleep Apnea Diagnosis Website 8/5/2011 4:21 PM
by tuftsepc
Welcome to the Tufts Evidence-based Practice Center website for "Future Research Needs in the Diagnosis of Obstructive Sleep Apnea". For more instructions on navigating the website click here: [Instructions to Stakeholders](#)

2. List of participants added to Reference Documents 8/5/2011 4:20 PM
by tuftsepc
Please see the document titled "6. List of Participants" in the Reference Documents section.

Video: Welcome to the Future Research Needs Project



Figure 3. Sample discussion thread: Diagnosis

Tufts CTSI Tufts Clinical and Translational Science Institute

Sleep Apnea Future Research Needs: Diagnosis

Advanced Search

Sleep Apnea Future Research Needs: Diagnosis

Sleep Apnea Future Research Needs: Diagnosis > Sleep Apnea Diagnosis FRN Topics > Diagnostic trials of questionnaire evaluating clinical outcomes

Sleep Apnea Diagnosis FRN Topics

View All Site Content

Home Page

Sleep Apnea Treatment Website

Contact Tufts Staff

Submit FRN Topic for Discussion

Sleep Apnea Diagnosis Topic Discussions

- 1. Association between use of clinical prediction rules and clinical outcomes
- 2. Association between use of questionnaires and clinical outcomes**
- 3. Head-to-head comparisons of portable monitors, questionnaires, and prediction rules
- 4. Randomized trials of phased testing
- 5. Routine (or selected) preoperative screening for sleep apnea
- 6. Defining the OSA syndrome
- 7. Value of having a sleep medicine specialist involved in the diagnosis of OSA
- 8. Indications (patient signs, symptoms, or other features) for appropriate home testing
- 9. Value of brain MRI in evaluating OSA patients
- 10. Value of scoring nasal flow limitation in recognizing mild OSA

Actions Settings View: **Threaded**

Threading Created By Replies

Edited: 7/19/2011 5:41 PM by [redacted] View Properties Reply

2. Association between use of questionnaires and clinical outcomes

Association between use of questionnaires and clinical outcomes

Rationale: Most studies evaluating questionnaires are focused on tests of accuracy like sensitivity and specificity. The most clinically useful evaluation of questionnaires would be studies examining whether use of the test resulted in improved clinical outcomes.

Instructions: Please comment on this topic for future research in treatment/diagnosis of sleep apnea. Please consider 1) the topic's importance; 2) whether the topic has already been sufficiently researched; 3) the feasibility of conducting research on this topic; and 4) its potential impact.

Posted: 8/10/2011 1:06 PM by tuftsepc View Properties Reply

Do people agree with the rationale? Are the questionnaires etc. more clinically useful to assess risk of bad clinical outcomes (eg, death, MI, diabetes, stroke) [ie, risk assessment] or to predict the likelihood that PSG will diagnose OSA [ie, value as use for pre-testing/phased testing]?

Show Quoted Messages

Posted: 8/12/2011 6:25 PM by [redacted] View Properties Reply

Because we are a rural based organization, access to PSG is limited and where available quite costly (little competition out here). We are also handicapped by fewer primary care providers than in urban settings so the performance of a self-administered questionnaire or an approach that lends itself to remote care (i.e., telemedicine) would be of special interest to us. Given that there is insufficient evidence linking the performance of questionnaires (Berlin, STOP, STOP-Bang, etc.) vs. PSG we would like to see more work done to determine if the accuracy of questionnaires can be relied upon as a cost-effective self-administered screening tool. We are willing to trade off slightly lower test accuracy for a gain in cost and consumer-friendliness domains. The lack of evidence linking questionnaires to clinical outcomes is, of course, the most valuable question but it would seem premature to research that relationship at this stage until there is better evidence that questionnaires are indeed a good (high +/- pred. values, etc.) screening tool (vs. PSG).

Show Quoted Messages

Posted: 8/15/2011 9:16 AM by tuftsepc View Properties Reply

Each FRN topic was uploaded as a separate “discussion board.” These were structured to allow stakeholders to comment on both the original FRN topic as well as other stakeholder comments (Figure 3). For their discussions, stakeholders were asked to consider four dimensions of need related to the proposed topic (see Appendix A and the Approach to Prioritization section, below). Stakeholder participation was monitored and discussions were moderated daily by Tufts EPC staff to ensure appropriateness and relevance of all comments. The discussion boards were initially scheduled to be open for a 2-week period. During the open period, stakeholders were regularly encouraged (two to four times per week) to contribute to the discussion boards as well as submit additional topics for discussion. Stakeholders were encouraged to go through all topic areas, and provide comments and feedback on all topics on the Web site.

We planned to use the SharePoint Web site to nominate FRN topics using a process similar to topic submission described above. This strategy was revised due to limited stakeholder engagement with the SharePoint Web site (as described in the Results section).

At the conclusion of the project, all online discussions and email communications were archived for transparency purposes.

Approach to Prioritization

The stakeholders were asked to consider four dimensions of need related to the proposed topic. These four dimensions come from the AHRQ Effective Health Care Program Selection Criteria (Appendix A). These dimensions and the Effective Health Care Program guidance on them were described in detail in the lead up to topic submission, discussion and selection of FRN topics.

- Importance
- Desirability of Research/Duplication
- Feasibility
- Potential Impact

EPC program staff evaluated a fifth dimension, Appropriateness, after submission of initial FRN topics.

After the close of the online discussion, the stakeholders were asked to identify up to 10 FRN topics that were of highest priority and that met the Effective Health Care Program selection criteria. The original plan was to conduct this nomination step on the Web site, but as described in the Results section, nomination was conducted by email and individual phone calls. After nomination, the EPC grouped similar topics into overarching topics and edited the names of the title for clarity and consistency. Based on the stakeholder nomination, the EPC categorized the overarching topics into four groups:

- High-priority FRN topics: Clearly of interest to stakeholders (based on high levels of nominations). A consensus of stakeholders expressed that these topics are of high priority. We aimed for about five topics, but we used natural breaks in the rankings of the topics, rather than strictly defining the numbers of topics in this category. For these topics, the background, stakeholder discussion, and study design considerations are fully elaborated in the Results section.
- Second-tier FRN priority topics: Of interest to a substantial number of stakeholders, but lacking a consensus that these were high-priority topics. For these topics, only a summary of the stakeholder discussion is presented.
- Other topics: Of relatively little interest to stakeholders. These include the remaining topics (see the next category for exceptions) that were nominated by few if any

stakeholders. Topics in this group are listed in the Results section without further discussion.

- Does not meet the Effective Health Care Program Appropriateness criteria. The EPC will move any topics that do not meet these minimum requirements for inclusion in the FRN report into this category for potential further discussion. These topics are not explicitly listed in the Results.

Approach to Research Question Development and Considerations for Potential Research Designs

For each high-priority FRN topic, we considered the range of study designs that would best address the topic. We did this taking into account the PICOD criteria (Population, Intervention, Comparator, Outcomes, study Design). For each topic, we described our assumptions about the most appropriate PICOD criteria, in particular describing the advantages and disadvantages of various potential research designs. We specifically considered the feasibility of the research questions focusing on potential sample size, time, and recruitment issues. For selected topics, we consulted with our local domain expert, who was also a stakeholder.

To determine candidate study designs, the feasibility of the study designs, and sample size calculations, we followed the structure laid out in the Future Research Needs document Framework for Considering Study Designs for Future Research Needs.⁴

Briefly, candidate study designs will differ across types of FRNs. Effectiveness or efficacy of treatments can be most definitively addressed in randomized controlled trials (RCTs), and secondarily in well-conducted nonrandomized comparative observational studies. In contrast, eliciting patient preferences can be meaningfully performed with nonexperimental designs (e.g., in a survey or focus group). Furthermore, observational studies may be most appropriate to enhance generalizability and determine effectiveness, as opposed to efficacy alone. Each final FRN topic was assessed as to context of the research question and a determination as to whether evaluation of efficacy or effectiveness is of greater need. This informs the choice of study design.

When a simple RCT (or diagnostic test study) was deemed to be an appropriate study design to address a FRN topic, we performed sample size calculations using standard formulae for a two-sided chi-squared test at the 0.05 level of significance. We assumed an allocation ratio of 1:1, no loss to followup, no crossover between treatments, and no sequential monitoring. For studies with dichotomous outcomes, we determined a range of reasonable control rates (event rates in the comparator arm) and relative effects (risk ratios) and based on these calculated a range of scenarios and minimum sample sizes. For studies with continuous outcomes, we range of reasonable mean differences in effect size (e.g., hours of sleep) between arm and standard deviations of the differences and based on these calculated a range of scenarios and minimum sample sizes. For continuous outcomes we also a conservative range of equivalence (the range between the smallest mean difference that would be clinically significant and the largest difference that could be reasonably expected). Where possible, these assumed values were derived from the existing evidence. All power calculations were set at 90 percent power to detect a significant result. We report all our assumptions.

Results

Composition of Stakeholder Panel

Through our multipronged approach to compiling the stakeholder panel, we enlisted a total of 27 stakeholders, of whom 21 were on the diagnosis panel. Table 2 summarizes the number and types of stakeholders on the panel.

Table 2. Stakeholders panel for diagnosis of sleep apnea

Category	Subcategory	No. of stakeholders
Patients and the public	Patient advocates or caregivers	4
	Current patients	
	Transportation sector employers	
Providers	Hospital administrator	6
	Clinicians - Primary care	
	Clinicians – Pulmonary care	
	Clinicians – Dental care	
	Clinicians – Psychiatry	
	Clinicians – Sleep medicine	
Purchasers	Private employers – Insured or Self-insured	2
	Public purchasers - e.g., Veterans Administration, Federal Employees Health Benefits Plan	
Payers	Private insurers	3
	Medicaid (at the State level)	
Policymakers	Agency for Healthcare Research and Quality	4
	Federal agencies	
	Professional organizations, Guideline developers	
Principal investigators	Clinical research	2
	Health services/policy research	
TOTAL		21

Stakeholder Participation and Required Methods Modifications

As noted above, this was the first instance of the Tufts Evidence-based Practice Center (EPC) producing a Future Research Needs (FRN) document with a large number of stakeholders. As such, the process was to a large extent a test of the success of the specific protocol we envisaged. Here we describe the substantive changes that we had to make to the protocol methods.

Overall, the use of the Sharepoint Web site was only moderately successful. The majority of stakeholders were slow to sign on to the Web site, were slow to begin reviewing the ongoing discussions, participated only minimally in discussions, and did not offer new FRN topics. Furthermore, a small number of participants were not able to log in to the Sharepoint Web site because of technical difficulties.

Nineteen FRN topics were discussed and prioritized; the EPC added five of these based on the Comparative Effectiveness Review (CER) Future Research section; stakeholders added eight online (or via email to the EPC); the EPC added six after a teleconference with the consumer advocates (see below).

The discussion period started 12 days after we received signed disclosure of interest forms from almost all the stakeholders (with the exception of the patient stakeholder who we were delayed in recruiting). All participants were contacted 5 days before the online discussion board was made available. We informed them of the impending start of the online discussions as

confirmed that their email addresses would be available to them during the course of the discussion. The original plan was for the stakeholder discussion to last 2 weeks; in reality, the discussion period was extended to 27 days to allow further stakeholder involvement. Four of 21 stakeholders signed in to the Web site within 2 days. The median time until stakeholders signed in was 7 days. Seven stakeholders did not log in to the online discussion site at all. The last stakeholder signed in on day 23 of the discussion. Across topics, only five stakeholders started discussions; seven stakeholders participated in the online discussions (added any comments). Not including comments or questions the EPC added, for the 19 topics discussed online, the median number of comments by stakeholders was 2, ranging from 0 (for five topics) to 8. After the discussion period ended, 16 of 21 stakeholders participated in topic nomination (which was conducted by email instead of on the Web site, as originally planned; see below).

To improve participation, after 1 week of discussion, we invited all stakeholders (from both panels) to participate in teleconferences where the EPC reviewed the materials that had been sent to them by email and that were available on the Web site, including the goals of the project, the stakeholders' responsibilities, the main criteria for selecting and discussing FRN topics, and a review of how to use the Sharepoint Web site. Two calls were scheduled, for which 13 of 27 stakeholders joined (from both diagnosis and treatment panels). To increase participation, we also sent numerous email reminders, offered to have stakeholders email us their comments that we would add to the discussion board, and also answered phone queries on technical issues related to the Web site. Furthermore, because several stakeholders could not easily access the Web site (primarily because of overseas travel), we compiled plain-text versions of all the FRN topics and discussions, which we placed directly into the text of emails (not as attachments). This was emailed to all stakeholders.

Near the true end of the discussion period, we noted that the consumer stakeholders (the patient, the patient advocate, and the representative from the transportation industry) had not participated in the discussion. Upon reviewing the discussions to date, we thought that they might have been too scientifically technical for the lay stakeholders. We contacted them directly, had our suspicions confirmed, and organized a separate teleconference for the three of them (with the EPC). This 1.5-hour teleconference produced six new topics. The EPC organized the discussion into FRN topics, summarized the separate discussions, sent the summary to the lay stakeholders, and then uploaded the topics and discussions to the Web site. A summary of the call was also emailed to all stakeholders. No stakeholders added further to these topic discussions.

After the discussion period was closed, the EPC decided that it was not worthwhile to attempt to use the Web site to have stakeholders nominate the topics. Instead this was done by email. The final list of topics and the text of the discussions were emailed to stakeholders. The order of the topics was randomized once, instead of separate randomization for each stakeholder, since the randomization process was too time consuming. (We chose to maintain links between the list of topics and the topic discussions, rather than manually reorder each stakeholder list.) To prevent procrastination and further delay, we asked stakeholders to nominate the topics within 4 days, by Friday, with the expectation that responses would be in by Monday. Eight of 21 stakeholders sent in the topic nominations by Monday. It was likely that some of the delay in response was due to a weather-related phenomenon (Hurricane Irene) so we extended the timeline and sent out a reminder email after 1 week. Six of the remaining 13 stakeholders sent in their nominations after receiving the email. Subsequently, followup phone calls were made to the remaining stakeholders to solicit their nominations. The final list of nominated topics was

received the following Monday. Topics were nominated by 16 of 21 stakeholders. Stakeholders were asked to nominate up to five topics; they prioritized between two and five topics each.

Research Needs

Based on the methods described above, we organized the FRN topics into three categories (see Methods section): High-priority FRN topics; second-tier priority FRN topics; and other FRN topics. Topics that did not meet minimum requirements for inclusion are not presented. The FRN topics are as listed in Table 3.

Table 3. List of future research needs topics

High Priority Future Research Needs Topics
1. Age- and gender-specific criteria for defining the OSA syndrome (and abnormal breathing)
2. Effect of routine (or selected) preoperative screening for sleep apnea
3. Cost-effectiveness analysis of a management strategy (diagnosis of symptomatic or high-risk patients through treatments of patients diagnosed with OSA), specifically for patients with mild to moderate disease severity
4. Value of having a sleep medicine specialist involved in the diagnosis of OSA (in addition to or instead of a nonspecialist)
5. What is the prognostic accuracy of clinical prediction rules (CPRs) to predict clinical outcomes?
Second-Tier Future Research Needs Topics
6. Indications (patient signs, symptoms, or other features) for appropriate home testing
7. Diagnostic approaches to OSA in obese and nonobese patients
Other Future Research Needs Topics
8. Can PSG be skipped in making the diagnosis of sleep apnea?
9. What are the financial barriers to access to diagnosis?
10. Head-to-head comparisons of portable monitors, questionnaires, and prediction rules
11. Association between use of questionnaires and clinical outcomes
12. What are the available, objectively-measured predictors of sleep apnea diagnosis?
13. What is consumer willingness-to-pay for screening, to identify consumer preferences for strategies to diagnose sleep apnea?
14. Value of scoring nasal flow limitation in recognizing mild OSA
15. Value of brain MRI in evaluating OSA patients
16. Randomized trials of phased testing
17. Value of using 4-phase rhinomanometry in recognition of patients with high nasal resistance and OSA
18. Diagnostic approach to OSA in micrognathia and retrognathia

High-Priority Future Research Needs Topics

High-Priority Future Research Needs Topic 1

Age- and gender-specific criteria for defining the OSA syndrome (and abnormal breathing)

Background

This FRN topic represents two topics that the stakeholders nominated and discussed separately. After nomination of topics by the stakeholder panel, the topics were combined into one overarching topic.

This overarching topic was addressed in general terms in the Introduction of the CER, but it was not a Key Question in the report. A Technology Assessment preceding the CER provided in-depth discussions on the following issues concerning the diagnosis of abnormal breathing or obstructive sleep apnea (OSA).⁵ Specifically, the Technology Assessment analyzed whether laboratory-based polysomnography (PSG) is a “gold standard” for measuring abnormal breathing

when establishing a diagnosis of OSA. It concluded that most experts consider laboratory-based PSG as the reference method to identify people with apnea-hypopnea index (AHI) suggestive of OSA. However, this does not mean that facility-based PSG is an error-free “gold standard” for measuring abnormal breathing, or that facility-based PSG measurements of breathing are generally sufficient for defining the OSA syndrome. Issues include the questionable reliability of PSG measurements, the variability in AHI scoring criteria, and that AHI does not correlate well with the intensity of the symptoms in patients who are believed to have OSA. Thus, the definition of the OSA syndrome requires information beyond the currently used measurements of breathing. This additional information includes signs and symptoms, and differentiation from other conditions that affect sleep. The lack of a “gold standard” for the diagnosis of OSA is reflected in the design of published studies included in the Technology Assessment and in the CER, where many different thresholds of AHI, ranging from 5 to 40 events/hr, have been used as suggestive of OSA in different studies. None of these studies used age- or gender-specific criteria.

Using AHI scores of ≥ 5 , ≥ 10 , and ≥ 15 events/hr to estimate the prevalence of sleep-disordered breathing in the general population showed that men had a higher prevalence of sleep apnea than women in middle-age groups (30 to 39, 40 to 49, and 50 to 60 years old) and at all AHI cutoffs.⁶ The prevalence of the condition among Medicare beneficiaries (people age 65 years or older) is believed to be higher than the aforementioned estimates among middle-age people.⁵

Stakeholder Discussion

The stakeholders stated that there is a need for a new definition of the OSA syndrome that identifies individuals with breathing abnormalities during sleep who have or are at risk of developing clinical sequelae because of their exposure to sleep-disordered breathing. This may require the use of other factors, including symptoms, biomarkers of stress induced by OSA, and markers of genetic risk. A better definition of OSA syndrome will more specifically target therapy at those who are at risk and avoid unnecessary intervention in those who are not. The stakeholders discussed that women generally have a lower AHI than men regardless of symptoms (e.g., snoring) for OSA. Thus, many women may not be diagnosed and remain untreated if using the same diagnostic criteria as men. Moreover, elderly (≥ 65 years old) may be another important subgroup of patients who need different diagnostic criteria for OSA. The stakeholders pointed out that there exist different AHI diagnostic criteria for school-age children. In the opinions of the stakeholders, studies to unveil age- and gender-specific criteria for defining OSA are feasible and would impact beneficially on treatment decisions and patient outcomes. It was further discussed that experts should be convened to establish standard diagnostic criteria and their corresponding epidemiologic implications for the different severity levels of OSA, and that these be used in future research. However, it was noted that this has already been done by the International Classification of Sleep Disorders (ICSD-2), but that the definition remains problematic.

Proposed Research Design

The aim of this topic is to identify age- and gender-specific criteria for defining OSA in individuals who are at increased health risk because of abnormal breathing during sleep. The aim is not to discuss the clinical utility of (yet undefined) age- and gender-specific diagnostic criteria.

To set up criteria for classifying people into categories, such as risk groups for abnormal breathing or OSA, the ideal approach has two stages. The first would be to discover the form of the relationship between specific risk-group criteria in the general population and long-term clinical outcomes. Once the form of the relationship is known, the second stage is to define (age- and sex-specific) thresholds for identifying those who are at increased health risk, that is, patients with the OSA syndrome. In principle, this could be done with (very large) epidemiological cohorts of people who are not receiving treatment for OSA. Obviously, it would be very challenging to design, fund and run such an epidemiological study. Further, a study that evaluates the relationships between AHI (or other PSG measures) and (untreated) clinical outcomes would have to be very long-term and would be unethical for symptomatic patients.

A study that merely describes the distribution of the measurements of breathing (e.g., AHI) in the general population, stratified by sex and age groups, would not conclusively address the challenge of defining OSA as a syndrome. First, currently used measurements of breathing during sleep (including AHI) are probably not sufficient for identifying those at increased health risk (i.e., those with the OSA syndrome); additional information such as signs and symptoms is necessary. Secondly, it does not measure any health-related outcomes that would help deduce which AHI measurements are “not normal”; it would be circular to use AHI (or other PSG measurements) both as a predictor of risk and also as an outcome. Third, it is also not sufficient to use symptoms suggestive of OSA as outcomes, because it is known that AHI scores (or other PSG measures) correlate poorly with symptoms. To establish criteria to predict symptoms would de facto exclude asymptomatic OSA; furthermore a PSG is not needed to determine symptoms. Ideally, one would like to evaluate hard clinical outcomes, such as mortality, cardiovascular disease, and similar endpoints.

Based on the considerations above and for simplicity, we suggest a prospective cohort study of people selected from the general population. The aim would be to associate measurements of breathing with a battery of short-term pathophysiological measurements that distinguish people whose breathing patterns have immediate physiological impact, from those whose breathing patterns do not have measurable functional sequelae.

Prospective Cohort Study

Value of Study Design

Prospective longitudinal cohort studies are the most informative studies to assess predictors of natural history outcomes (e.g., age and sex as criteria for OSA). A well-designed prospective study will be less biased than a retrospective database analysis. Randomized trials of interventions generally do not provide better data to assess predictors of natural history outcomes.

To set up age- and gender-specific diagnostic criteria for abnormal breathing or OSA, we propose a prospective cohort study of people selected from the general population. All subjects would receive facility-based PSG for sleep-disordered breathing. The primary outcome of interest would be the functional outcomes of the effects of OSA on physiology (“pathophysiological measurements”). Possible functional outcomes include cerebral hypoxia or vasoconstriction. Sleep medicine domain experts would have to choose proxy outcomes that are measurable in the short term (e.g., overnight during a PSG), are highly likely to be due to the apnea, and are highly likely to be the proximate cause of long-term clinical outcomes. Multivariable analyses that control for potential confounders, such as comorbidity and body

mass index, would be performed to evaluate the relationship between AHI scores (or other PSG measures) and functional outcomes by sex and age groups. Ideally, all age- and gender-specific criteria derived from such studies would be verified in multiple independent cohorts of patients by independent researchers. It should be noted that the definitive definition of OSA will remain unclear because whether the functional outcomes are good proxy markers for clinical outcomes will likely remain unknown.

With long-term followup, the proposed cohort study can also provide natural history data to be used to assess whether patient characteristics (e.g., age, sex, race/ethnicity), in addition to disease severity, play a role in determining long-term clinical outcomes. Any subsequent treatments and additional diagnostic testing could also be recorded as these data would inform the clinical utility of the age- and gender-specific diagnostic criteria.

Resource Use, Size, and Duration

Such a study would require the enrollment of several thousand people. The population-based studies that aimed to estimate the prevalence of sleep-disordered breathing included at least 1,000 individuals. Multivariable analyses would need larger sample sizes than univariable analyses to reach the same statistical power. The study would not impose additional diagnostic testing (beyond the initial PSG) or treatments compared to usual care; therefore, the main nonadministrative resources would be gathering followup outcome data on all patients. Ideally, the study should follow participants longitudinally with multiple time points for outcome assessments. The followup frequencies should be more often within the first year and then annually.

Ability To Recruit

It should be relatively easy to recruit patients into a prospective cohort study. The study would be mostly observational in nature, so the only major added burden to patients would be having an initial PSG and then providing followup outcome data.

Ethical Issues

Since no diagnostic decisions or treatments are being imposed on patients, there are no ethical issues directly related to such a study.

High Priority Future Research Needs Topic 2

Effect of routine (or selected) preoperative screening for sleep apnea

Background

Key Question 3 of the CER asked: What is the effect of preoperative screening for sleep apnea on surgical outcomes? The CER found insufficient evidence to address this topic. Two poor-quality studies assessed the effect of preoperative screening for sleep apnea on surgical outcomes. One study found no significant differences in outcomes between patients undergoing bariatric surgery who had mandatory PSG or PSG based on clinical parameters. The second study found that general surgery patients willing to undergo preoperative PSG were more likely to have perioperative complications, particularly cardiopulmonary complications, possibly suggesting that patients willing to undergo PSG are more ill than other patients. No trials have addressed the value of routine (or selected) preoperative screening for OSA.

Stakeholder Discussion

Most of the discussion centered on the anecdotal nature of the evidence regarding the perioperative risks in patients with undiagnosed OSA. The stakeholders talked about the concern among anesthesiologists and surgeons about the patients who do not present with high-risk characteristics (e.g., a nonobese patient without obvious airway obstruction), as well as those patients who either fail to wean from the ventilator or have respiratory failure after discharge. These patients were felt to belong to the group of patients undergoing thoracic or abdominal procedures. The respiratory complication rate in patients with OSA was not thought to be high, and it was felt that very large study sample sizes would be needed to evaluate any strategy to identify high-risk patients. The effectiveness of perioperative care was discussed as well as strategies for mitigating risks such as involving experienced anesthesiologists rather than a nurse or resident anesthetist. Furthermore, stakeholders discussed whether the question applies to all surgical cases, only elective cases, or high-risk surgeries (e.g., upper airway surgery, procedures that require patients to remain supine for a period of time postoperatively). Outcomes of interest listed by one stakeholder include time to weaning, days in the intensive care unit, length of hospital stay, and complications such as line infections, ventilator-associated pneumonia, and urinary tract infections. It was also noted that studies may be difficult because they could not be blinded and that behaviors across an institution may change just by virtue of how patients are managed in the trial.

Proposed Study Designs

To address the overarching FRN topic and its subcomponents, different study designs are reasonable and would address different aspects of this topic.

Randomized Controlled Trials

Value of Study Design

The CER found no trials of preoperative screening. A randomized comparison of screening for OSA with extended anesthesia care when appropriate versus no screening with routine anesthesia care would provide information for effectiveness of the screening protocol. However, the sample size calculations below highlight the large number of participants who would need to be recruited for such a randomized controlled trial (RCT). Therefore, the design of choice would likely be a multicenter cluster-randomized trial, where whole centers would be randomized to screen patients or not. The advantage of using a cluster-randomized trial is that there is a minimal risk of protocol deviation if active OSA screening is performed as part of a center's clinical protocols. There would also be little risk of cross-contamination of the culture of screening for OSA to patients who had been randomized to not receive screening. It is also likely that recruitment and randomization will be logistically easier if it is done at the center level, rather than at the patient level within a center.

Resource Use, Size, and Duration

The primary analysis of interest in such a study would be the comparative effectiveness of screening on postoperative morbidity and mortality. The most common instruments that could be used as part of the preoperative bedside screening include questionnaires—Berlin, STOP (Snoring, Tiredness during daytime, Observed apnea, and high blood Pressure), the STOP-Bang

(STOP with body mass index, age, neck circumference, and sex variables), the American Society of Anesthesiologists (ASA) screening checklist, and CPRs—based on a combination of clinical measurements, questionnaire answers, airway measurements and demographic variables. The duration of the intervention would be the perioperative period, when followup is routinely done as part of normal clinical practice.

Due to its rarer occurrence, a mortality outcome would necessitate a larger sample size than for an outcome like respiratory complications. However, with the advances in postoperative care, the rates of even complications after surgery have declined considerably, such that sample sizes would need to be large even for these outcomes. To estimate minimum sample sizes (power) we used the data from one of the studies in the report that compared mandatory PSG screening with screening based on clinical risk factors.⁷ We used standard formulae for a two-sided equivalence test. We used the postoperative intensive care unit admission rate reported in the study as an outcome. We set the range of relative risk reduction between 0.6 and 0.8, as the study reported a relative risk of 0.6. We set the range of event rates in the control arm from 0.01 to 0.1, as the control arm in the study reported an event rate of 0.05. Assuming a power of 90 percent, the required total sample size (1:1 ratio in a two-arm RCT) for each value of relative risk and the control rate is presented in Figure 4. With the very low rates of postoperative complications, the required sample sizes range from 2,000 to 100,000 participants. If we reduce the power to 80 percent, the range of sample sizes is still large, ranging from 1,500 to 70,000 participants as the relative risk and control rate increases. This highlights the tradeoff that one would have to make between having sufficient power to address the effectiveness of preoperative screening and the large amount of resources that it would need to undertake such a trial. If a convincing argument can be made that a continuous outcome could be an adequate proxy outcome for postoperative complications, then it is likely that a smaller sample size would be needed for adequate power for this outcome. However, it is currently unclear than any continuous outcome would be a convincing proxy.

Ability To Recruit

The sampling population is large, as there are thousands of patients who undergo inpatient and outpatient surgery. Patients are usually concerned about intraoperative and postoperative complications and would be interested in using management strategies to mitigate them. If centers, rather than patients, were randomized, it may be possible to consider such a trial to be an examination of an interview procedure, which may allow for a waiver of informed consent. Loss to followup should be rare during the period of hospitalization, but may be of concern for posthospitalization followup unless an adequate system were in place to follow patients.

Ethical Issues

The primary ethical issue (beyond the standard ethical issues in conducting an RCT) pertains to whether there is equipoise. It is not known whether performing active preoperative screening for OSA would result in improved perisurgical care and would thus lead to decreased postoperative morbidity and mortality. The use of randomized study designs allows us to answer whether active screening results in improved morbidity and mortality. Thus, there is clinical equipoise.

Observational Studies

Value of Study Design

Observational studies, such as those that exist, are likely to be fundamentally flawed. In normal clinical practice, preoperative patients who undergo testing for OSA will always be greatly different than unscreened patients. It is unlikely that any amount of statistical adjustment, including propensity scores, could overcome the major clinical differences between the groups of patients. One possible exception to this may be an observational comparison of two or more similar hospitals or clinics where routine screening for OSA is done in some, but not others. However, even in this situation it is unlikely that a convincing argument could be made that any differences in outcomes are primarily due to the use of OSA screening, as opposed to a whole host of other differences that are likely to coexist.

Resource Use, Size, and Duration

Retrospective data could be gathered quickly and easily by chart review or similar approaches. Prospective cohorts would take somewhat more time and effort, but would still be less resource intensive than a trial. Since the outcomes ascertainment is potentially completed within a short time frame (within a few weeks) after the operation, data could be collected fairly rapidly, whether retrospective or prospective. However, the validity of the data collected using a flawed study design detracts from the benefit gained by decreased resource utilization.

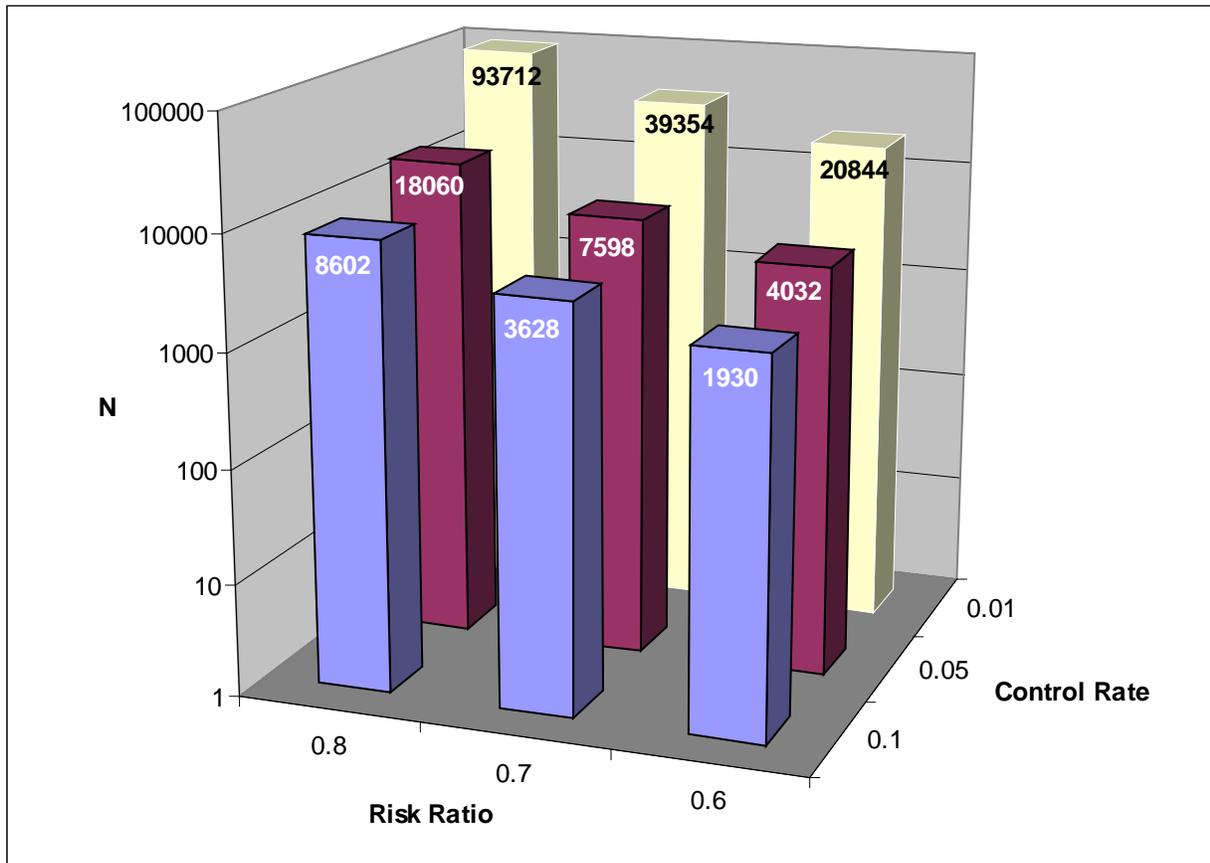
Ability To Recruit

Patient recruitment should be straightforward and relatively simple, as there would be little added burden for them by entering a study, beyond a formalized preoperative screening protocol, all of which are reasonable parts of normal patient care.

Ethical Issues

The primary ethical issue involved would be related to recruiting patients and expending resources in a study that is unlikely to provide convincing results.

Figure 4. Sample size calculations for an RCT of screening versus no screening for postoperative complications with varying estimates of risk ratio and control rates



High-Priority Future Research Needs Topic 3

Cost-effectiveness analysis of a management strategy (diagnosis of symptomatic or high-risk patients through treatments of patients diagnosed with OSA), specifically for patients with mild to moderate disease severity

Background

The topic as proposed covers both diagnosis and treatment strategies for a complete cost-effectiveness analysis. Since diagnosis and treatment FRN topics are being dealt with separately, here we focus primarily on the diagnosis strategy component of the cost-effectiveness analysis. A discussion of a cost-effectiveness analysis of treatment strategies can be found in the companion report.³ It would likely be of value for future cost-effectiveness analyses to incorporate both phases of OSA management (diagnosis and treatment). Cost-effectiveness analysis allows the comparison of different interventions on similar benefit, cost and utility scales. If clinical trial data are available on the treatments and populations of interest, benefit and cost estimates have both internal and external validity.

Stakeholder Discussion

The objective of this FRN project is to establish better evidence about the costs and benefits of alternative management strategies for individual patients with mild to moderate disease severity. Ultimately, the goal is to develop evidence about care management from efficient testing of high-risk individuals and accurate and low-cost diagnosis of OSA. Issues that should be considered in a cost-effectiveness model include the patient-related outcomes, including measures of functional status such as productivity (both absenteeism and presenteeism—working in spite of illness, with resulting poor work performance), quality of life, and work safety. The feasibility of such an analysis will depend on combining clinical trial results with health insurance data on health care utilization, employer data on absenteeism, and potentially auto insurance data on motor vehicle accidents in a large cohort of patients with OSA. Other issues that should be considered are the patients' costs for using OSA treatments. An example of such a cost, which would not be included in most analyses, was the costs to a long-haul truck driver who requires extra battery backup since a continuous positive airway pressure (CPAP) device can drain the truck's battery overnight.

Proposed Study Designs

Systematic Review

Cost-effectiveness analyses were not addressed by the CER. Conducting a systematic review may be the first step to ascertain the level of existing evidence.

Cost–Benefit Analysis

On the basis of the best evidence identified in the CER, a quality-adjusted cost-benefit analysis is recommended, comparing the incremental costs and benefits of different diagnostic strategies to each other. Benefit, utility, and cost estimates may be derived from previous clinical study data, where available, and from observational data where trial data are not available. These estimates should include not only the standard clinical outcomes, but also work-related, accident, and quality of life outcomes. In the absence of cost estimates, charges may be derived from administrative data and adjusted cost-to charge ratios. Out-of-pocket patient costs should also be included.

Analytic Approach

Preference should be given to a Markov-chain, discreet events analysis. The proper outcome measure is quality-adjusted life-years (QALYs) gained. Future QALYs and costs should be discounted over the followup period. Adjustment should be made for major outcomes and adverse events, using rate and utility estimates. Probabilistic sensitivity analyses should be conducted on estimated outcome rates, utilities and cost estimates. A societal perspective should be assumed in the main analysis. The patient perspective should be assumed in a subanalysis, given high out-of-pocket costs for diagnostic strategies and substantial concerns about burden on the patient.

Diagnostic Strategies

Diagnostic strategies should be tested both with and without the involvement of a sleep medicine specialist, where possible. Phased treatment combinations should also be considered. A short list of treatment strategies to be compared could include:

- A standard diagnosis procedure (diagnosis using a questionnaire in the primary care setting)
- Other diagnostic algorithms and CPRs
- Portable monitors
- Scoring nasal flow limitation
- Four-phase rhinomanometry
- Brain magnetic resonance imaging

Resource Use, Size, and Duration

Because a cost-effectiveness analysis can draw from previously collected data, the cost, size, and duration of such studies can be limited. However, as discussed by the stakeholders, it may be challenging to gather all the relevant data.

Ethical, Legal, and Social Issues

No new data collection is proposed, and therefore, the direct risk to patients is minimal.

High-Priority Future Research Needs Topic 4

Value of having a sleep medicine specialist involved in the diagnosis of OSA (in addition to or instead of a nonspecialist)

Background

Increasingly patients with OSA are being diagnosed by primary-care providers. This occurs because some sleep centers allow direct referral for sleep studies and then leave it to the primary-care physician to provide followup. Furthermore, home studies have allowed the primary-care physician to bypass the sleep center altogether. National companies have started marketing the use of home studies to primary-care providers. The CER did not address the effect of having different specialists involved in care. To ascertain the level of existing evidence, it may be prudent to conduct an initial systematic review.

Since most of the scientific studies involving diagnosis of OSA are carried out in highly specialized sleep centers with patients under the supervision of a sleep medicine specialist, it is unknown how effective an approach that does not include input from a sleep specialist would be. Of note, a recent survey found that physicians who specialized in certain fields associated with sleep medicine (pulmonary medicine, neurology, and psychiatry) had different practice patterns compared to other physicians with respect to prescriptions, education, and adherence to CPAP therapy.⁸ With the high prevalence of undiagnosed OSA, full testing of all individuals at high risk for OSA would likely overwhelm the capacity of sleep centers and sleep medicine specialists.

Stakeholder Suggestions

The stakeholders discussed that OSA is sufficiently common that it is unrealistic that millions of patients are going to be able to see a small number of specialists. Requiring the involvement of a sleep specialist may also present a barrier to care in many settings and would likely increase costs. The stakeholders who participated in the discussion were skeptical about the added value of including sleep medicine specialists in the initial diagnosis. However, it was noted that many primary care physicians do not feel adequately trained or comfortable with

management of these patients. The stakeholders were interested in comparisons between diagnosis (or more broadly management) by primary care physicians alone, primary care physicians in concert with a specialist, and specialists alone.

Proposed Study Designs

Three different study designs are proposed to explore the differences in outcomes when patients are initially diagnosed by sleep specialists or by physicians who are not sleep specialists.

Systematic Review

The CER did not address the effect of having sleep specialists involved in care. Conducting a systematic review may be the first step to ascertain the level of existing evidence.

Analyzing Claims Data

Value of Study Design

One approach to ascertaining cost differences between patients that are diagnosed by specialists as opposed to those diagnosed by nonspecialists is to analyze claims data provided by an insurance provider or a health care system. This approach would require a dataset that included a mix of patients diagnosed by both approaches. One possible dataset would be deidentified Medicare and/or Medicaid claims data. This approach would be able to provide a large sample of patients from diverse geographic locations, along with accurate data on cost of care, including physician, diagnostic testing, and treatment costs, as well as outcomes such as time to diagnosis.

An important limitation of this approach is that the analysis is not controlled. Thus, there may be factors associated with seeing a sleep specialist for initial diagnoses that are associated with outcomes that are unrelated to the care provided by the sleep specialist. Despite this, analyzing claims data from different geographical areas would provide a solid basis for further prospectively designed trials.

Resource Use, Size, and Duration

Analysis of an existing dataset does not require the substantial costs and time spent with patient recruitment and followup.

Ethical Issues

Data would be deidentified and retrospectively analyzed. Thus, there would be no ethical issues in analyzing this data.

Post Hoc Analysis of Existing Trials

Value of Study Design

Given the breadth of diagnostic and treatment studies analyzed in the existing systematic review of OSA, it could be possible to review these studies to analyze any available information on diagnosis by a sleep specialist versus diagnosis by a nonspecialist. Studies may present this information in a variety of different ways. First, a given study may include only patients that were diagnosed by sleep apnea specialists. Presumably, this would be the majority of studies. Second, a study may include all or some patients from diagnosis by a nonspecialist, or from

diagnosis by a team that includes both specialists and nonspecialists. Outcomes from these studies may be compared, if information is given on cost-of-care and/or sleep-related outcomes such as AHI and sleepiness measures (e.g., Epworth Sleepiness Scale). Finally, a study may provide subgroup information on patients that were diagnosed by a sleep specialist or nonspecialist, providing a more direct comparison of outcomes. However, it is likely that to adequately address the topic, data additional to what is included in published articles would need to be gathered. This would depend on the willingness of primary investigators to share their unpublished data.

Resource use, Size, and Duration

Reviewing studies included in the existing systematic review would not require costs associated with a prospective study, but could involve a substantial amount of time for reanalyzing a large number of papers.

Ethical Issues

Since data would be collected from existing studies, there would be no ethical issues involved.

Survey of Providers

Value of Study Design

A cross-sectional survey of providers who diagnose patients would provide an in-depth view of issues associated with using a sleep specialist in the diagnosis of OSA. This survey could provide information on the variety and magnitude of issues faced by nonspecialists who may not have the in-depth knowledge of sleep specialists. In addition, this survey could capture specific information related to cost-of-care. This could include information such as number of patients seen per hour as well as factors associated with differential cost-of-care like geographical location and specific field of specialty (e.g., pulmonary medicine, neurology, and psychiatry). The questions asked in the survey would need to be phrased in a neutral fashion to avoid biasing the survey respondents' answer to favor sleep specialists or nonspecialists with respect to any outcome measure or practice pattern.

Resource Use, Size, and Duration

The size of the survey would likely be limited by participation rates. Providers could be contacted who are part of a sleep medicine association to capture specialists as well as primary care providers with an interest in sleep apnea issues. To include primary care providers who are not specifically interested in sleep apnea issues, an additional database of providers would have to be used. As the study is cross-sectional in nature, the duration of the study is not an issue.

Ethical Issues

There are no ethical issues of concern.

High Priority Future Research Needs Topic 5

What is the prognostic accuracy of clinical prediction rules (CPRs) to predict clinical outcomes?

Background

The CER found a low strength of evidence among seven studies that some clinical prediction rules (CPRs) may be useful in the prediction of a diagnosis of OSA. Ten different CPRs have been described. Nine CPRs have been used for the prediction of a diagnosis of OSA (using different criteria). Among those, the oropharyngeal morphometric model gave near perfect discrimination (area under the curve [AUC] = 0.996) to predict the diagnosis of OSA, and the pulmonary function data model had 100 percent sensitivity with 84 percent specificity to predict diagnosis of OSA. The remaining models reported lower sensitivities and specificities. Each model was deemed useful to predict the diagnoses of OSA by the individual study authors. However, while all the models were internally validated, external validation of these CPRs has not been conducted in the vast majority of the studies. None of the studies examined whether use of the CPR resulted in improved clinical outcomes.

Stakeholder Discussion

While the stakeholders considered the topic to be important, most of the discussion centered on the various factors that influence the diagnostic accuracy of CPRs, namely the setting (intercountry differences), level of care (primary, secondary, tertiary care) or source of data for initial screening assessment (self-report by patient, reporting by patient's bed partner, or evaluation by health care provider).

Proposed Study Designs

We clarify that the aim of the topic is to evaluate the “prognostic” accuracy of existing CPRs, not to discuss the development of novel CPRs to predict OSA diagnosis by PSG. Specifically, the aim is to evaluate whether CPRs can determine who will experience a clinical outcome in the future. Different CPR thresholds will correspond to different counts of patients falling into different risk (or prognostic) groups based on clinical outcomes; the number of clinical events in the risk categories identified by the thresholds will vary depending on the actual value of the threshold used. Arguably, the most appropriate study design to address the prognostic accuracy of CPRs is a prospective observational study.

Prospective Observational Study

Value of Study Design

CPRs will be used to assign a “risk category” to each patient according to their likelihood of experiencing a future clinical outcome. Well-performing CPRs would be those that have high discriminatory ability and retain good calibration when they are evaluated in various settings. Discrimination pertains to whether the CPR is able to distinguish between categories of increasing risk of the experiencing the outcome. In the simple case of only two risk categories (low vs. high), the (predictive) sensitivity and the (predictive) specificity of a CPR would be a way to measure its discriminatory ability. Calibration pertains to whether the proportions

predicted by the CPR in each risk category are “close” to the actual observed proportions. High calibration implies good discrimination, but good discrimination does not imply good calibration. For most clinical purposes, CPRs would be most informative if they have good calibration characteristics. However, for the purposes of this exposition, it is simpler to provide a sample size analysis for metrics that are related to the discriminatory ability of a CPR.

Prospective observational studies are best suited to assess the prognostic value of a CPR because they can study multiple risk factors of interest as well as clinical outcomes in a general patient population in whom the CPR will eventually be used, ensuring applicability and external validity of the results. The CPR would be applied to participants upon their entry to the cohort, and they would be followed to ascertain whether they experience an outcome or not. Another design is a nested case-control study, where all patients in a defined cohort who experience an event are designated as “cases” and are matched to a collection of “controls” (i.e., cohort participants who did not experience the outcome). A major advantage of a nested case-control study is that it allows an adequately powered post hoc analysis of prospectively collected data where the event rate is relatively rare.

The CPRs can be tested for their prognostic value for a range of outcomes, including sleep measures (which are quick and easy to collect), symptoms (e.g., sleepiness scales), clinical outcomes (e.g., mortality, cardiovascular disease, and diabetes; which would require decades-worth of followup), or their intermediate outcomes (e.g., blood pressure or measures of glucose and insulin homeostasis). Of note, the stakeholders were principally interested in the prognostic ability of CPRs on long-term clinical outcomes (instead of their surrogates).

Resource Use, Size, and Duration

If suitable databases are immediately available, one can perform a case-control study in a relatively short time, as there is no need to wait for followup. However, it may be difficult to reconstruct the CPR or identify people who match the setting of interest from a retrospective database, and therefore prospective collection of data may be unavoidable.

Prospective cohorts would take more time and effort compared to a case-control study built on an existing database. For example, when assessing incident cardiovascular disease or diabetes mellitus, the timeframe could be months or years. When assessing mortality outcomes, the timeframe could very well extend to decades. As mentioned above, already published studies of CPRs used short-term surrogate or intermediate markers in place of long-term clinical outcomes (for example, change in blood pressure for cardiovascular disease and change in fasting blood glucose for glucose intolerance).

Because of the substantial resources necessary for a prospective cohort study or a nested case-control study, it is probably not practical to design a study whose sole purpose is to assess the prognostic ability of CPRs. Instead, it would be preferable to incorporate the assessment of the prognostic performance of CPRs into a prospective cohort study in which assessment of prognostic performance will be one of several aims.

We performed exploratory power calculations to obtain approximate estimates of sample sizes for a nested case-control study or a cohort study. For simplicity, we assume that the goal is to enroll enough participants to estimate the discriminatory ability of a CPR (i.e., its sensitivity and specificity) with reasonable precision, operationally defined as a confidence interval length (upper minus lower bound) of no more than 0.10 (or 10 percent) for the sample estimates of both

sensitivity and specificity: * For example, if the sample sensitivity is 50 percent, the confidence interval would be no wider than 45 percent to 55 percent. † The sample size calculations also depend on the prevalence of OSA in the setting of interest. Here, we assume that we are interested in applying the CPR to the general population, where the prevalence of OSA is approximately 10 percent.⁹

Table 4 shows hypothetical sample size calculations for three scenarios of CPR sensitivity and specificity pairs. The first scenario is a case of low sensitivity (44 percent) and high specificity (85 percent), which was reported with a CPR based on age, sex, body mass index, reported snoring, and reported cessation of breathing during sleep.¹⁰ The second scenario is where the sensitivity and specificity were over 80 percent, seen with a CPR based on a multivariable apnea prediction questionnaire score and oximetry (sensitivity 83 percent, specificity 95 percent).¹¹ The third scenario includes a high sensitivity (84 percent) but low specificity (39 percent) seen with a CPR based on a 24-item questionnaire and clinical characteristics.¹²

Ability To Recruit

Patient recruitment should be straightforward and relatively simple, as patients are interested in knowing whether they have a condition that is known to cause complications and is associated with chronic disease outcomes.

Ethical Issues

The major ethical issue in this study design is the risk of a false negative report, resulting in denial of treatment to those patients who truly need it.

Table 4. Sample size estimates for various clinical scenarios of sensitivity and specificity with an assumed prevalence of 10 percent and an allowed variance of 10 percent in both sensitivity and specificity

Scenario	Sensitivity (%)	Specificity (%)	N _{Sensitivity} [*]	N _{Specificity} [†]	N _{case-control} [‡]	N _{cohort} [‡]
Low sensitivity High specificity	44	85	379	196	575	3790
High sensitivity High specificity	83	95	217	73	290	2170
High sensitivity Low specificity	84	39	207	366	573	2070

* N needed to estimate sensitivity or specificity with a confidence interval that has length at most 0.10.

† This refers to a case control-study that would be conducted using existing records. We assume 1:1 ratio between cases and controls.

‡ N needed for Cohort Study: The number of people who have to be analyzed to ensure that we will have the necessary totals for estimating both sensitivity and specificity with the desirable precision. For example in scenario 1, since the prevalence is 10 percent, the number of people that you would have to screen to recruit 379 confirmed OSA patients is $379 \times 10 = 3790$ and the number of people that you would have to screen to recruit 196 people without OSA is 245. Since the larger number needed to screen for the given sensitivity would also satisfy the number needed to screen for the given specificity, the final number needed to screen is 3790.

* Ideally, what constitutes “reasonable precision” would be defined based on additional information, and can differ according to the envisioned role of the CPR in clinical practice. The decision to use a length of 0.10 for the confidence interval is for illustration.

† The boundaries of the confidence intervals are not symmetric for proportions other than 0.50, but the desired length of the confidence interval is operationally defined to be 0.10.

Second-Tier Future Research Needs Topics

Second-Tier Future Research Needs Topic 6

Indications (patient signs, symptoms, or other features) for appropriate home testing

Background

This topic was not directly addressed in the CER, except as the indications may have been incorporated into CPRs.

Stakeholder Discussion

The rationale for this topic discusses how the typical patient being evaluated for sleep apnea does not meet the criteria for eligibility into most trials, namely a high probability of OSA and a lack of significant factors that increase risk of other forms of sleep disordered breathing. Of interest is whether the home testing diagnostic studies are generalizable to more typical, lower risk patients.

Second-Tier Future Research Needs Topic 7

Diagnostic approaches to OSA in obese and nonobese patients

Background

The subquestion to Key Question 1 of the CER addressed how different available tests compare in different subgroups of patients, including by body weight. The evidence was found to be insufficient to address this subquestion.

Stakeholder Discussion

The discussion revolved around the need to improve the diagnosis of, and refocus attention on, OSA patients who are not overweight, but instead have other causes for OSA.

Other Future Research Needs Topics

Other Future Research Needs Topic 8

Can PSG be skipped in making the diagnosis of sleep apnea?

Other Future Research Needs Topic 9

What are the financial barriers to access to diagnosis?

Other Future Research Needs Topic 10

Head-to-head comparisons of portable monitors, questionnaires, and prediction rules

Other Future Research Needs Topic 11

Association between use of questionnaires and clinical outcomes

Other Future Research Needs Topic 12

What are the available, objectively measured predictors of sleep apnea diagnosis?

Other Future Research Needs Topic 13

What is consumer willingness to pay for screening, to identify consumer preferences for strategies to diagnose sleep apnea?

Other Future Research Needs Topic 14

Value of scoring nasal flow limitation in recognizing mild OSA

Other Future Research Needs Topic 15

Value of brain MRI in evaluating OSA patients

Other Future Research Needs Topic 16

Randomized trials of phased testing

Other Future Research Needs Topic 17

Value of using 4-phase rhinomanometry in recognition of patients with high nasal resistance and OSA

This topic was not considered to be high priority by any stakeholder.

Other Future Research Needs Topic 18

Diagnostic approach to OSA in micrognathia and retrognathia

A stakeholder sent this topic to the EPC after the panel discussion ended.

Discussion

Challenges in Stakeholder Involvement

We implemented a Web-based discussion board in preference over a series of teleconferences because of what we believed would be advantages of the online approach. Namely, greater flexibility for stakeholders to participate in the discussion at times convenient to them, including nonworking hours; a platform that would allow everyone to have a more equal voice, where discussion would not be led by the stakeholder who verbally dominated; full participation by all stakeholders in all discussions, not just those that occurred during calls they were able to attend; a full record of all discussions, without the need to summarize verbal discussions which inevitably leads to omissions and other errors; less time expenditure by stakeholders who would not be asked to sit through numerous 60-90 minute phone calls; considerably less resource expenditure by Evidence-based Practice Center (EPC) staff, where each teleconference hour translates into 7 person-hours of resources; and less opportunity for “multitasking” during discussion (e.g., answering emails while participating in teleconferences).

However, we faced several challenges when implementing the SharePoint Web site discussion board. Perhaps the greatest challenge was the low participation rate of stakeholders during most stages of the project. When the discussion period was announced, few participants logged onto the Web site within the first week. Because of the low log-in rate, combined with the low number of page views for most stakeholders, the discussion period had to be extended. Despite extending the discussion period, low participation rates persisted. To increase participation, we sent several email reminders to stakeholders. After the discussion period was closed, participation rates remained low during the nomination period, despite simplification of the process by asking for nomination by email. We subsequently telephoned remaining stakeholders, which increased total participation. Overall, although SharePoint offered the convenience of asynchronous collaboration it seemed to lack appropriate incentives to engage stakeholders. Notably, the seemingly egalitarian Web-based workspace failed to connect patient advocates with the other stakeholder representatives. Assembling patient advocates independently via group teleconference proved more effective, but limited the integration of patient perspective into the SharePoint discussion forums.

Apparent barriers to using the Sharepoint Web site more fully included technical trouble logging in (e.g., while traveling); apparent reluctance by stakeholders to devote unscheduled time to sign in to the Web site, review the discussions and comment; hesitancy commenting in the online discussion by stakeholders without a technical background; and possibly a lack of engagement and interaction from communicating by text instead of verbally. The large number of topics may have been too great a burden for stakeholders to read at a single sitting. In addition, the timing of the project during the summer months may have limited participation because of vacation schedules. A large part of the problem of poor participation may have been simply that the Web-based approach did not force people to schedule a time to participate (as they would for a teleconference), thus there may have been a lack of perceived urgency. Furthermore, for almost all topics, the comments were so sparse that there was little sense of a discussion for stakeholders to participate in.

Several possible solutions exist for these problems. One potential approach which may improve discussion on an online discussion board would be to increase the depth of topic discussion by reducing the number of stakeholders. This could be accomplished by purposely

limiting the number of participants to those who show the ability and proven willingness to participate in discussions at early stages. For example, a run-in phase could allow inclusion of stakeholders, who could be given the opportunity to offer new topics and start discussions, but then further discussion would continue with only the interested stakeholders. In effect, this is what occurred, but an a priori plan to limit the number of participating stakeholders could help to make the process more explicit and possibly shorten the time period required for discussion. A potential downside of this approach is that it could be seen as more explicitly biased since certain stakeholders would be dropped from further participation. Also, the practical problem of how to cordially disinvite stakeholders from further participation would have to be overcome.

Alternatively, the EPC could revert to the more standard and basic approach of conducting a series of teleconferences, where all available stakeholders could join any or all calls. The teleconferences could have either a loose agenda where any new or old topic could be addressed, or a more structured agenda where each call would focus on a general category of Future Research Needs (FRN) topics. We used this approach near the end of the discussion period, when we noted that the consumer stakeholders (the patient, the patient advocate, and the representative from the transportation industry) had not participated in the discussion. We organized a separate teleconference for the three of them and a 1.5-hour teleconference produced six new topics. The EPC would need to work to ensure that all participating stakeholders are given multiple opportunities to fully express their thoughts. The EPC would also have to accurately summarize the calls in an unbiased fashion, a nontrivial task. The summaries would have to be in a form that could be relatively easily used by the stakeholders.

A similar approach could be to use focus groups to allow a full, simultaneous discussion of each topic. Focus groups could be formed by either collecting stakeholders with similar backgrounds together to focus on specific topics or by randomizing stakeholders to focus groups in order to obtain a balance of viewpoints. These approaches could allow more in-depth discussion of focused topics. However, potential downsides include lessening the participation of less vocal or less topic-knowledgeable individuals and setting up separate subgroup discussions that not all stakeholders have equal access to. This approach may be somewhat simpler to organize and would require less time commitment by stakeholders than multiple teleconferences for all stakeholders.

An approach that combined both teleconferences with online discussions may be able to combine the advantages of the two discussion types, without many of the disadvantages of each. However, this approach is likely to be more time-consuming than the original timeline planned (though possibly less time-consuming than the actual discussion period for the current project). One variation could be an initial gathering of FRN topics and rationale text online, followed by teleconferences to further discuss each topic. The teleconferences would be summarized by the EPC staff and uploaded to the Web site. Further discussion would then be encouraged on line. Clearly, numerous other similar approaches are possible.

Regardless of which type of discussion were held, there may be advantages to limiting the size of the stakeholder panel that could improve discussion. We took the approach of being expansive to include as many voices as feasible. However, in future, it may be better to limit the panel to a single member of each stakeholder category. Attempts should be made to choose stakeholders who show enthusiasm in joining the panel and participating in discussions. However, unfortunately, across all the stakeholders we did not find a strong correlation between the level of enthusiasm upon invitation and the degree of participation in discussions.

The stakeholder discussion identifying FRN topics on diagnosis of obstructive sleep apnea (OSA) highlights the vast field of research that could potentially be done in this area. Three out of the five nominated topics fell outside the purview of the questions addressed in the CER. The scope of the original CER was focused on the use of portable tools and screening, phased testing and preoperative screening, and the relationships of OSA indices to clinical outcomes. Notably, the stakeholders considered a range of topics broader than the CER's Key Questions regarding diagnosis of OSA to be important for future research.

In summary, the online Web site discussion of numerous topics by a large stakeholder panel was only moderately successful. Various approaches to improve participation and discussion are possible, including increased use of teleconferences, restricting the size and members of the stakeholder panel, and other approaches. More experience with different approaches is needed.

References

1. Balk EM, Moorthy D, Obadan NO, Patal K, Ip S, Chung M, Bannuru RR, Kitsios GD, Sen S, Iovin RC, Gaylor JM, D'Ambrosio C, and Lau J. Diagnosis and Treatment of Obstructive Sleep Apnea in Adults. Comparative Effectiveness Review No. 32. (Prepared by Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-1). AHRQ Publication No. 11-EHC052-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2011.
2. Agency for Healthcare Research and Quality. Future Research Needs Projects. 2010. www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayProduct&productID=521. Accessed September 6, 2011.
3. Balk EM, Chung M, Moorthy D, Chan JA, Patel K, Concannon TW, Ratichek SJ, Chang LKW. Diagnosis of Obstructive Sleep Apnea: Future Research Needs. Future Research Needs Paper No. 11. (Prepared by Tufts Evidence-based Practice Center under Contract No. 290-2007-10055 I.) AHRQ Publication No. 12-EHC031-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2012. www.effectivehealthcare.gov/reports/final.cfm.
4. Agency for Healthcare Research and Quality. Framework for Considering Study Designs for Future Research Needs (FRN). Draft Comparative Effectiveness Review. Rockville, MD: Agency for Healthcare Research and Quality; September 2011. www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=774.
5. Trikalinos TA, Ip S, Raman G, et al. Home Diagnosis of Obstructive Sleep Apnea-Hypopnea Syndrome. Technology Assessment. (Prepared by the Tufts-New England Medical Center Evidence-based Practice Center.) Rockville, MD: Agency for Healthcare Research and Quality; 2007. www.cms.gov/determinationprocess/downloads/id48TA.pdf.
6. Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. *N Engl J Med.* 1993;328:1230-1235.
7. Hallowell PT, Stellato TA, Petrozzi MC, et al. Eliminating respiratory intensive care unit stay after gastric bypass surgery.[erratum appears in *Surgery.* 2008 Feb;143(2):301]. *Surgery.* 2007;142:608-612.
8. Parthasarathy S, Haynes PL, Budhiraja R, et al. A national survey of the effect of sleep medicine specialists and American Academy of Sleep Medicine Accreditation on management of obstructive sleep apnea.[see comment]. *Journal of Clinical Sleep Medicine.* 2006;2:133-142.
9. Young T, Finn L, Peppard PE, et al. Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort. *Sleep.* 2008;31:1071-1078.
10. Rodsutti J, Hensley M, Thakkinstian A, et al. A clinical decision rule to prioritize polysomnography in patients with suspected sleep apnea. *Sleep.* 2004;27:694-699.
11. Gurubhagavatula I, Maislin G, Pack AI. An algorithm to stratify sleep apnea risk in a sleep disorders clinic population. *American Journal of Respiratory & Critical Care Medicine.* 2001;164:t-9.
12. Rowley JA, Aboussouan LS, Badr MS. The use of clinical prediction formulas in the evaluation of obstructive sleep apnea. *Sleep.* 2000;23:929-938.

Acronyms

AHI	Apnea-hypopnea index
AHRQ	Agency for Healthcare Research and Quality
CER	Comparative Effectiveness Review (on OSA, by Tufts EPC) ^{1]}
CPAP	Continuous positive airway pressure (device)
CPR	Clinical prediction rule
EPC	Evidence-based Practice Center
FRN	Future research need(s)
OSA	Obstructive sleep apnea
PSG	Polysomnography
QALY	Quality-adjusted life years
RCT	Randomized controlled trial

Appendix A. AHRQ's Effective Health Care Program Selection Criteria for New Research

1. Appropriateness	1a. Represents a health care drug, intervention, device, technology, or health care system/setting available (or soon to be available) in the United States
	1b. Relevant to 1013 enrollees (Medicare, Medicaid, S-CHIP, other federal health care programs)
	1c. Represents one of the priority conditions designated by the Department of Health and Human Services (DHHS)
2. Importance	2a. Represents a significant disease burden; large proportion or priority population
	2b. Is of high public interest; affects health care decisionmaking, outcomes, or costs for a large proportion of the US population or for a priority population in particular
	2c. Was nominated/strongly supported by one or more stakeholder groups
	2d. Represents important uncertainty for decisionmakers
	2e. Incorporates issues around both clinical benefits and potential clinical harms
	2f. Represents important variation in clinical care, or controversy in what constitutes appropriate clinical care
	2g. Represents high costs due to common use, to high unit costs, or to high associated costs to consumers, to patients, to health care systems, or to payers
3. Desirability of New Research/Duplication	3. Would not be redundant (i.e., the proposed new research is not sufficiently researched by AHRQ or others, considering both completed and in-process research)
4. Potential Impact	4a. Potential for significant health impact: - To improve health outcomes - To reduce significant variation in clinical practices known to be related to quality of care - To reduce unnecessary burden on those with health care problems
	4b. Potential for significant economic impact: - To reduce unnecessary or excessive costs
	4c. Potential for change: - The proposed topic exists within a clinical, consumer, or policymaking context that is amenable to evidence-based change - A product from the EHC program could be an appropriate vehicle
	4d. Potential risk from inaction: - Unintended harms from lack of prioritization of a nominated topic
	4e. Addresses inequities, vulnerable populations (including issues for patient subgroups)
	4f. Addresses a topic that has clear implications for resolving important dilemmas in health and health care decisions made by one or more stakeholder groups
	5. Capacity
	5b. Utilizes existing AHRQ resources or builds desired additional research capacity or decisional support for the EHC Program
	5c. Costs associated with the likely study design are reasonable considering limited program resources