Considerations and Evidence for an ADHD Outcome Measure

Donna Woods, EdM, PhD; Mark Wolraich, MD; Karen Pierce, MD; Lindsay DiMarco, MPH; Nicole Muller, BS; Ramesh Sachdeva, MD, PhD, MBA

From the Center for Healthcare Studies, Northwestern University, Feinberg School of Medicine, Chicago, Ill (Dr Woods, Ms DiMarco, and Ms Muller); University of Oklahoma Health Sciences Center, Oklahoma City, Okla (Dr Wolraich); Northwestern University Feinberg School of Medicine, Chicago, Ill (Dr Pierce); American Academy of Pediatrics, Elk Grove Village, Ill (Dr Sachdeva); and Medical College of Wisconsin, Milwaukee, Wis (Dr Sachdeva)

The authors declare that they have no conflict of interest.

Address correspondence to Donna Woods, EdM, PhD, Center for Healthcare Studies, Northwestern University, 710 N Lake Shore Dr, 10th Floor, Chicago, IL 60611 (e-mail: woods@northwestern.edu).

Received for publication October 31, 2013; accepted June 18, 2014.

OBJECTIVE: The 2011 American Academy of Pediatrics attention-deficit/hyperactivity disorder (ADHD) guideline emphasizes monitoring and measuring outcomes of children diagnosed with ADHD; however, recommendations for how to measure improvement are less clear. A long-term goal was to develop an outcome measure that assesses the quality of care for children with ADHD. As a first step in that process, we conducted a literature synthesis on the efficacy and effectiveness of guideline-recommended ADHD treatments on patient outcomes.

METHODS: A literature search was conducted in PubMed according to PRISMA protocol and using MeSH terms. US Preventive Services Task Force (USPSTF) criteria were used to assess the level of evidence. Studies of interest were published after 2002 and assessed prospective ADHD improvement using recommended ADHD treatments.

RESULTS: The systematic review resulted in 35 studies. According to USPSTF criteria, included studies were level I (n = 24), level II-1 (n = 1), and level II-2 (n = 10) and were rated as good (n = 20) or fair (n = 15). DSM-criteria-based rating scales were used most frequently to measure ADHD treatment outcomes. All included treatments resulted in ADHD improvement. Regardless of outcome measure, tool, or treatment type, symptom reduction and improvement were relatively large, with mean percentage reductions ranging from 20% to 86% on ADHD-Behavior Rating Scales scales, with only 1 study with <25% reduction. Effect sizes ranged from 0.15 to 4.57.

CONCLUSIONS: On the basis of this literature review, a consistent pattern of improvement in pediatric ADHD patients’ core symptoms emerged across studies, study designs, and recommended treatment approaches. This evidence supports the notion that an improvement of core symptoms within 1 year could satisfy the requirements of an effective outcome measure, which should be further investigated.

KEYWORDS: attention-deficit/hyperactivity disorder; mental health; outcome measures; pediatrics; quality of health care

ACADEMIC PEDIATRICS 2014;14:S54–S60

ATTENTION-DEFICIT/HYPERACTIVITY DISORDER (ADHD) is a commonly identified neurobehavioral disorder characterized by inattention, hyperactivity, impulsivity, and related functional impairment.1 ADHD has a multidimensional effect on daily functioning, incurring significant costs, attributable to greater health care needs and lost productivity. According to the Centers for Disease Control and Prevention (CDC),2 5 million children (9%) between 3 and 17 years of age have been diagnosed with ADHD. The percentage of children 4 to 17 years of age with a parent-reported ADHD diagnosis increased by 22% between 2003 and 2007 and rates of ADHD diagnosis increased an average 5.5% per year from 2003 to 2007.3 High rates of ADHD were noted among children of racial and ethnic minority groups (15.7%) and children covered by Medicaid (13.6%).

In 2011, the American Academy of Pediatrics (AAP) published an evidence-based guideline4 for pediatricians summarizing the empirical literature and making recommendations regarding the provision of effective ADHD diagnosis and care. Accurately diagnosing ADHD requires a comprehensive evaluation, which includes completion of standardized objective rating scales based on the criteria of the current Diagnostic and Statistical Manual of Mental Disorders1,5 by parents and teachers to assess children for the presence of ADHD symptoms in at least 2 settings. A critical recommendation of the 2011 ADHD clinical guideline is systematic follow-up, which includes a recommendation for collection of follow-up parent and teacher rating scales to quantitatively reassess symptoms, function, and adverse effects or risks related to medication treatment.

As part of a larger initiative, established through the Children’s Health Insurance Program Reauthorization Act of 2009 (CHIPRA), the Agency for Healthcare Research and Quality (AHRQ) led an effort called the Pediatric Quality Measures Program (PQMP) and was charged with the development of evidence-based pediatric...
quality measures to assess a broad range of pediatric health care quality topics. These measures are intended for voluntary use by Medicaid and CHIP programs, as well as other public and private payers, plans, and providers. AHRQ funded the Pediatric Measures Center of Excellence (PMCoE) and assigned development of pediatric quality measures for ADHD care.

In Donabedian and Attwood’s seminal 1963 work, quality domains and measures were categorized as either measures of structure, process, or outcome. Hospitals and other health care providers continue to express concerns that patient health outcomes have many determinants, some beyond the control of health care providers. Thus, they argue, it is unfair and invalid to hold providers accountable for overall patient health outcomes. Although this perspective dominated the field of quality measurement for decades, there is increasing interest in using patient health outcomes as quality indicators as patients are concerned most about what happens to their health and well-being.

The Initial Child Core Measure Set currently includes an ADHD process measure, which assesses only whether a patient with ADHD had follow-up visits and does not measure elements of visit quality or patient’s ADHD symptom improvement. As a result, our long-term goal was to develop an outcome measure that assesses the quality of care for children with ADHD. As a first step in that process, we conducted a literature synthesis on the efficacy and effectiveness of guideline-recommended treatments for ADHD on patient outcomes. In this report, we present the results from this literature review and discuss the implications of these findings for the construction of a new outcome measure of the quality of care for ADHD.

The goal of ADHD treatment is the improvement of symptoms and functional impairment; and gaps in performance in pediatric ADHD care have been established in the literature.

When a treatment has a demonstrated efficacy, defined as the probability of benefit to individuals in a defined population from a medical innovation applied for a given medical problem under ideal condition of use, measuring care outcomes can provide an important indicator of the effectiveness of the care delivered, or benefit of a medical treatment under ordinary conditions provided by the average practitioner, for the typical patient.

An outcome-based measure must be relevant, meaningful, measurable, and actionable by the clinician and/or family for quality improvement purposes and must be important, reliable, valid, and feasible to be useful as a health care quality measure for accountability. Although symptom reduction has been used extensively to measure outcomes in randomized control trials, there has been no determination of the validity of using these evaluations as outcomes to measure the quality of ADHD treatment and follow-up care in actual practices. Although improvement in symptoms and functional impairment is the goal of any ADHD treatment, validated measures of improvement in functional impairment are not available or used. The determination of improvement in functional impairment is a clinical decision and no standardized assessments or measures are applied in practice. The aim of this study was to inform the long-term goal of the development of an outcome measure to assess the quality of care for children with ADHD. The purpose of this study was to conduct a literature synthesis on the efficacy and effectiveness of evidence-based, recommended treatments on patient outcomes, present these results, and discuss the implications of these findings for the construction of a new outcome measure of ADHD care quality, including the implications for the development of quality measures suitable for Meaningful Use Stage III.

METHODS

This literature review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol.

DATA SOURCE AND SEARCHES

First we developed the strategy for the search. Next, on February 14, 2014, a search was conducted of primary studies, systematic reviews, and meta-analyses in the PubMed database pairing the Medical Subject Headings (MeSH) term “Attention Deficit Disorder with Hyperactivity” with each of the following: “Quality Improvement,” “Treatment Outcome,” “Outcome and Process Assessment,” and “Outcome Assessment.”

STUDY SELECTION

We included peer-reviewed studies, regardless of design, if the study assessed prospective ADHD pediatric patient improvement given 2011 ADHD clinical guideline–recommended treatments. Eligible study settings included outpatient clinics, psychiatric institutes, and the patient’s home. Studies were included if they were published January 1, 2002, after the previous 2001 AAP ADHD guideline was published, to February 14, 2014. Because the focus was US care quality, studies conducted in health systems outside the United States were excluded. As minimum criteria for study quality, we excluded studies that were rated poor in quality according to the USPSTF criteria or that had a very low N (N<15).

Three researchers independently screened the article titles. Abstracts were then screened by the 3 researchers and full-text studies assessed for eligibility. Each study was reviewed by 2 reviewers. Studies were excluded if they did not meet the criteria presented in Online Appendix 1. We obtained full-text articles when either investigator found the abstract potentially eligible. Discrepancies in the determination of the eligibility of full-text articles were resolved by consensus according to the study criteria.

DATA EXTRACTION AND QUALITY ASSESSMENT

The level of evidence of studies was assessed according to the US Preventive Services Task Force (USPSTF), Hierarchy of Research Design Criteria (Online Appendix 2). Studies with an evidence grade below level II, rated as poor, or with a very small sample size (<15) were excluded.
Two investigators independently extracted data from each study using a standardized form (Online Appendix 3). Disagreements were resolved by consensus. Extracted elements included: Patient population, age, intervention, comparators, principal measures of ADHD outcomes, study timing, and study setting and results (percent improvement) — and were entered into a standardized table. Studies were organized according to ADHD treatment modality (Online Appendix 3).

Results from ADHD treatment (eg, percentage reduction in symptoms, effect size, etc.) were then arrayed according to outcome measure tool (Table). In addition, when reported in studies, effect sizes are included in the Table. When effect sizes were not directly reported but the studies included the appropriate elements to calculate effect size, Cohen’s d effect size was calculated by finding the difference in means and then dividing the difference by the standard deviation of the control group.

**RESULTS**

**DESCRIPTION OF STUDIES**

A total of 3363 studies were identified in the PubMed electronic search. After a title review, 3127 studies were eliminated according to the exclusion criteria, 456 of which were duplicates. Of the remaining 236 studies, 140 were excluded after abstract screening for criteria relevance, and US settings. After a full-text review, 13 articles were excluded because they were not conducted on treatment provided in the United States, 12 had a follow-up time of <30 days, 10 were editorials or commentaries, 9 were rated as “poor” by USPSTF criteria, 6 reported on treatments not included in the ADHD guideline, 6 were conducted in a laboratory setting, 3 had small sample sizes (<15), and 2 were unrelated to the study question (Figure).

Thirty-five studies met criteria to be included in the comprehensive literature review (Online Appendix 3). Twenty-four of the studies were rated level I (Online Appendix 2), with 17 of the studies considered good and 7 considered fair. One study was rated as a level II-1 and was considered fair, and 10 of the studies were rated as level II-2, with 3 considered good and 7 considered fair. Recommended treatments from the 2011 ADHD clinical guideline included: 1) medication therapies: methylphenidate (n=5), lisdexamfetamine dimesylate (n=2), atomoxetine (n=9), methylphenidate versus atomoxetine (n=2), clonidine extended release (n=1), guanfacine extended release (n=3); 2) behavior therapy (n=5); and 3) combined medication and behavior therapy (n=8). Study settings varied, including outpatient clinics, psychiatric institutes, and the patient’s home. Studies that included children with major comorbidities were excluded, and 6 of the 35 studies specifically excluded subjects with IQs of <70 or who were not performing academically at an age-appropriate level.

**STUDY RESULTS**

Online Appendix 3 presents the description and results of included studies by treatment. All included treatments resulted in ADHD improvement. Improvement in symptoms ranged from 20% to 86%, with only 1 study reporting <25% reduction.

**OUTCOME MEASURES OF ADHD TREATMENT**

Four types of principal measures of ADHD outcomes were used in included studies: 1) DSM-criteria-based rating scales, 2) Clinical Global Impression scales, 3) direct observation, and 4) other (n = 2). Most studies either solely relied on the DSM-criteria-based rating scales (n = 6, 17%) or used these in combination with 1 of the other 3 types of assessment instruments (n = 23, 66%). Overall, 86% of included studies used DSM-criteria-based rating scales to assess improvement. These assessments provide a global symptom score and subtype symptom scores. Of the 6 remaining studies that did not use DSM-criteria-based rating scales, 2 assessed symptom improvement using different methods. The Table synthesizes the range of improvement and available effect sizes by assessment instrument for the included studies.

**OUTCOME MEASURE TOOLS**

Although the included studies used many different outcome measure tools, core symptom reduction and improvement were relatively large regardless of which tool was used to assess improvement (Table). Mean percentage reductions ranged from 20% to 86% on ADHD-RS scales, with only 1 study reporting <25% reduction. Some studies reported improvement from initial and follow-up assessments on only teacher rating scales or only parent rating scales. Study effect sizes ranged from 0.15 to 4.57. There were no studies that assessed improvement in functional impairment.

**DISCUSSION**

This study represents the first comprehensive research synthesis of the literature related to ADHD improvement and outcomes to inform quality outcome measurement of current evidence-based pediatric ADHD care. All studies reviewed showed improvement and demonstrated that effective ADHD treatment and follow-up resulted in reduction of core symptom scores compared with the initial assessment scores, at the time of diagnosis (Table). Given the consistent improvement seen in the results across numerous studies, improvement in symptoms could provide the basis for the development of an outcome measure to assess the quality of care for pediatric patients diagnosed with ADHD.

All studies that assessed long-term results demonstrated that improvement in symptoms was maintained or improved over time if treatment continued after the stabilization of medication dosages. If follow-up is not effective and either ADHD medications or behavior therapy is stopped through attrition or lack of adherence, the treatment effect will stop and the symptom reduction initially seen will no longer be present. Reduction of symptoms as an outcome of care assesses not only effective treatment initiation, but also follow-up and continued ADHD treatment use. Although different behavior rating scales were
employed in the included studies, they all utilized the same symptoms that were rated with the same scale format.

The Clinical Global Impression—Improvement scale (CGI-I) was also used in numerous studies to assess improvement. The CGI-I could provide an assessment of improvement in functional impairment beyond symptom reduction; however, it has not been validated for this use and is not routinely used in clinical care. The CGI-I, as used in most studies, was global and not specific to ADHD. Changes are also likely to be slower using this tool.

The range of outcomes presented in this systematic review is largely based on results from carefully constructed research studies and therefore may overrepresent the magnitude of improvement achievable in the community. Epstein et al. reports on actual community-based care, and found at least a 25% decrease in symptom scores on parent and teacher DSM-criteria-based rating scales for 77% and 81% of patients, respectively, after 6 weeks of treatment initiation. Similar findings were reported in the community group of the MTA study, leading us to conclude that long-term improvement in average domain scores could address this.

The challenge of accurately diagnosing ADHD and differentiating between ADHD and other mental health conditions remains a known limitation. However, the recommendation to reassess the ADHD diagnosed patient

### LIMITATIONS

Most of the studies included in this systematic review were conducted in a research context, where medication was carefully titrated and patients received research level intensive follow-up care. Some included studies conducted in the community (normal care conditions), had less favorable findings as a result of issues with adherence and follow-up. However, care coordination and ADHD follow-up care are important components of the 2011 ADHD guideline recommendations to support treatment adherence and achieve treatment goals.

Practices differ in the volume of ADHD patients. Assessment of how sensitive this measure is to volume and the number of patients necessary for a stable estimate of the outcome and comparison of performance will be needed.

The table below presents the range of improvement observed in various outcome measure tools:

<table>
<thead>
<tr>
<th>Outcome Measure Tool</th>
<th>Improvement</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSM-IV-based scales</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADHD-RS (Parent, Teacher, and Investigator)</td>
<td>Mean score reduction: 9.6–28.6</td>
<td>19–26, 31, 32, 42, 43</td>
</tr>
<tr>
<td></td>
<td>Mean percentage reduction: 20%–60%</td>
<td>19–23, 25, 26, 30–32, 42–44</td>
</tr>
<tr>
<td></td>
<td>Effect sizes*: 0.15–4.09</td>
<td>19, 20, 22, 25, 27, 30, 32, 42–44, 45–47</td>
</tr>
<tr>
<td>Conners' Rating Scales (Parent and Teacher)</td>
<td>Reduction in mean score: 1.3–1.6</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Mean percentage reduction: 25%–86%</td>
<td>19, 21, 49</td>
</tr>
<tr>
<td></td>
<td>Effect sizes: 0.34–0.95</td>
<td>21, 46</td>
</tr>
<tr>
<td>SNAP (Parent and Teacher)</td>
<td>Proportion with score ≤1: 25%–68%</td>
<td>50, 51</td>
</tr>
<tr>
<td></td>
<td>Effect sizes: 0.22–1.60</td>
<td>24, 29, 36</td>
</tr>
<tr>
<td>Checklist of DSM Criteria</td>
<td>Reduction in mean number of symptoms: 7.5–8.7 (42%–48%)</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Effect size: 0.72</td>
<td>45</td>
</tr>
<tr>
<td>Disruptive Behavior Rating Scale</td>
<td>Mean score reduction: 4.7</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Effect size: 0.75</td>
<td>45</td>
</tr>
<tr>
<td>Global assessment scales</td>
<td>Percentage of patients rated as very much improved/much improved: 43%–99%</td>
<td>19, 23, 24, 27, 29–32</td>
</tr>
<tr>
<td>Clinical Global Impression—Improvement scale</td>
<td>Change in percentage of patients rated as normal to mildly ill: 59.8%–88%</td>
<td>19, 29, 30</td>
</tr>
<tr>
<td></td>
<td>Reduction in scores: 1.2–1.5</td>
<td>19, 43</td>
</tr>
<tr>
<td></td>
<td>Effect size: 4.57</td>
<td>32</td>
</tr>
<tr>
<td>Children’s and Parent’s Global Assessment Scales</td>
<td>Mean change in scores: 14–19</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Mean values: 2.7–3.4</td>
<td>19</td>
</tr>
</tbody>
</table>

*Effect sizes noted when reported or possible to calculate. Reduction in scores and percentage reduction are indicators of attention-deficit/hyperactivity disorder improvement.
significant performance gaps between evidence-based care and standard community-based care for pediatric patients diagnosed with ADHD. Current quality measures for ADHD follow-up care are imperfect. Developing a meaningful outcome measure requires balancing appropriate clinical expectations as well as the resulting implications for pediatric clinicians with the maximum benefit that is to be expected for pediatric patients. This systematic evidence review determined there is sufficient evidence to support measuring clinical outcomes in practice. An improvement in symptoms within a year could satisfy the requirements of an effective outcome measure for Meaningful Use and should be pursued for further research.

ACKNOWLEDGMENTS

We acknowledge the valuable contributions of the many experts and staff who made this report possible. We particularly recognize the PMCoE ADHD Expert Workgroup for their clinical expertise, national perspective, and leadership. We thank all of the PMCoE ADHD Expert Workgroup members who have so generously given their time and contributed their expertise to this valuable work. We acknowledge and express our gratitude for the extremely valuable comments and perspective offered by Denise Dougherty, PhD, Christopher Forrest, MD, PhD, and Kamila Mistry, PhD, MPH in the presentation of our findings and the editing of this manuscript. We would like to acknowledge the contribution of Raymond Kang in providing statistical analytic support. Supported in part by the US Department of Health and Human Services Agency for Healthcare Research and Quality and Centers for Medicare & Medicaid Services, CHIPRA Pediatric Quality Measures Program Centers of Excellence under grant number U18-HS-20498. The views expressed herein are those of the authors and do not necessarily represent that of the funding sources.

SUPPLEMENTARY DATA

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.acap.2014.06.011.

REFERENCES


