Risk Adjustment and Outcome Measures for Out-of-hospital Respiratory Distress

Samuel M. Keim, MD, Daniel W. Spaite, MD, Ronald F. Maio, DO, MS, Herbert G. Garrison, MD, MPH, Jeffrey S. Desmond, MD, Mary Ann Gregor, MHSA, Patricia J. O’Malley, MD, Ian G. Stiell, MD, MSc, C. Gene Cayten, MD, MPH, John L. Chew Jr., MS, Ellen J. MacKenzie, PhD, David R. Miller, MBA

Abstract

The purpose of the Emergency Medical Services Outcomes Project (EMSOP) is to develop a foundation and framework for outcomes research. In prior work, this group delineated the priority conditions, described conceptual models, suggested core and risk-adjustment measures potentially useful to emergency medical services research, and summarized out-of-hospital pain measurement. In this fifth article in the EMSOP series, the authors recommend specific risk-adjustment measures and outcome measures for use in out-of-hospital research on patients presenting with respiratory distress. The methodology included systematic literature searches and a structured review by an expert panel. The EMSOP group recommends use of pulse oximetry, peak expiratory flow rate, and the visual analog dyspnea scale as potential risk-adjustment measures and outcome measures for out-of-hospital research in patients with respiratory distress. Furthermore, using mortality as an outcome measure is also recommended. Future research is needed to alleviate the paucity of validated tools for out-of-hospital outcomes research. Key words: respiratory distress; outcomes; pulse oximetry; peak expiratory flow rate; visual analog dyspnea scale; mortality; out-of-hospital research; risk adjustment. ACADEMIC EMERGENCY MEDICINE 2004; 11:1074–1081.

This is the fifth article in a series reporting the results of the work of the Emergency Medical Services Outcomes Project (EMSOP). EMSOP was a five-year project funded by the National Highway Traffic Safety Administration with the aims to identify 1) conditions that take precedence in emergency medical services (EMS) outcomes research, 2) risk-adjustment measures (RAMs) for these priority conditions, and 3) outcome measures (OMs) for these priority conditions. Prior articles have described the initial investigators (RFM, HGG, DWS), consultants (JSD, CGC, JLC, DRM, EJM, PJO, IGS), expert panel members, general methods, and priority conditions; described conceptual models; suggested core and risk-adjustment measures potentially useful to EMS research; and summarized out-of-hospital pain measurement. In this article, measures that can be used for risk adjustment and outcomes evaluation for the out-of-hospital condition of respiratory distress are identified and discussed.

In a prior article, out-of-hospital conditions were rank-ordered by an expert panel based on condition frequency and expert opinion of the potential impact of EMS intervention on the outcome of that condition. The 42 individuals making up the expert panel included 23 physicians and 19 career EMS professionals. Their names, affiliations, and rationale for involvement are detailed in the prior work. In addition, the experts ranked the relevance of outcome categories (death, discomfort, disability, disease, destitution, and dissatisfaction) for each out-of-hospital condition. In this weighted priority ranking, respiratory distress ranked second among conditions for adults and third among conditions for children. An EMS condition was defined as an illness, injury, or combination of signs and symptoms that cause EMS activation. Implicit in this definition is that a single
EMS condition can encompass multiple diagnoses or diseases. Several specific disease states fall under the condition of respiratory distress, including but not limited to asthma, chronic obstructive pulmonary disease (COPD), and congestive heart failure (CHF). This produces one of the major challenges in identifying RAMs and OMs for EMS. This problem, however, is intrinsic to out-of-hospital care because patients present in the field with conditions, not diagnoses. Out-of-hospital care providers are, however, trained to make and act on condition assessments, not diagnosis. Additionally, the “impurity” introduced by the necessary mixing or grouping of multiple diagnoses in a “condition” has not prevented meaningful outcomes research when care has been taken to develop robust methods.7–13

METHODS

This study involved repeated explicit literature searches followed by consensus group reviews.

Definition of Terms.

Risk-adjustment Measure. A RAM is a variable that meaningfully reflects a patient’s characteristics and clinical attributes or otherwise affects or confounds a patient’s outcome in some way. RAM techniques often may also serve as OMs, for example, blood pressure measurements that are taken before (RAM) and after (OM) an intervention, although the preintervention technique would always be considered a RAM and the postintervention technique an OM. Previous work by the EMSOP group has described a conceptual framework for risk adjustment and outcomes evaluation in EMS research.2 Some RAMs will be similar across conditions, and a standard set of “core” EMS RAMs has been described.3 Others are explicit for specific conditions (e.g., Revised Trauma Score). RAMs are essential ingredients for investigations that target the effectiveness of EMS interventions related to dyspnea to minimize potentially confounding factors.

Outcome Measure. An OM is a variable that meaningfully reflects one or more of the outcome categories of death, discomfort, disability (functional impairment), disease, destitution (cost), and dissatisfaction (satisfaction).5,6 Some OMs may be applicable across conditions, whereas others are specific to a particular condition. Ideal OMs for out-of-hospital use would be easily and quickly applied, be applicable to all ages, and not require prolonged training or expensive, complicated, or bulky equipment.

For this study, RAMs and OMs were identified by a systematic literature search and a structured review of original research articles pertaining to each potential measure. Measures were evaluated by a method previously used to develop OMs in physical therapy by the Canadian Physiotherapy Association.14 After evaluation, each measure was discussed and a decision made to recommend or discard the measure. The methodology of the search and reviews, findings, and recommended measures were presented to and reviewed by a consultant group (JSD, CGC, JLC, DRM, EJM, PJO, IGS) and are shown in Figure 1.

Phase One.

Step 1: Literature Review Strategy. The initial phase consisted of a MEDLINE search of English-language articles from 1986 to 1996 by the primary investigators (RFM, HGG, DWS) using the Ovid (Ovid Technologies) search engine. An initial set of references was developed by combining the search for respiratory distress with a search for RAMs and OMs. MeSH search terms chosen by the investigators are shown in Figure 1.

Step 2: References Limited. This initial set of references was then limited to English-language articles that pertained to human subjects and were published in Abridged Index Medicus (AIM) journals. This created a final reference set of 2,836 references. The titles were reviewed by the investigators in a structured manner. References that focused on development or evaluation of a measure were included for further review. Studies that included the measures in clinical trials but did not specifically evaluate RAMs or OMs were not included. A study was included for further review if any single investigator chose it. A unanimous rejection of a title was required for the reference to be eliminated from further review. A total of 497 titles were selected for further review.

Step 3: Abstracts Reviewed. Abstracts of the selected references were obtained and reviewed. Each abstract was reviewed by all of the investigators. Papers not evaluating the feasibility, reliability, or validity of RAMs or OMs were excluded from the next review. An abstract required unanimous rejection by the investigators to be excluded from further evaluation. A total of 175 abstracts were selected for further review. For each abstract selected, the full-length article was obtained.

Step 4: Articles Reviewed and Sorted. Examination of these full-length papers resulted in 75 studies focused on the development or evaluation of a RAM or OM. Articles were then sorted into groups based on the measure addressed (e.g., dyspnea scales or measurements of pulmonary function). An implicit structured review of each group of articles pertaining to each measure was conducted by a single investigator using the guidelines from “Physical Rehabilitation Outcome Measures” published by the Canadian Physiotherapy Association.14 The attributes evaluated for each measure included time taken to complete the measure,
cost and training, scaling, reliability, and validity. These guidelines were modified to include feasibility of use in the out-of-hospital setting. Reviews were conducted independently and documented in a standardized fashion. References listed in the articles that pertained to the development of RAMs or OMs were obtained and reviewed.

**Figure 1.** Methodology for identifying risk-adjustment measures (RAMs) and outcome measures (OMs). CPA = Canadian Physiotherapy Association.

**Step 5: Group Presentations and Consensus.** The written individual reviews were presented to the entire group of investigators and consultants. Each investigator orally presented the results of his or her review following the modified Canadian Physiotherapy Association guidelines and made recommendations regarding the appropriateness of the measure for out-
of-hospital outcomes research. After each presentation, a discussion ensued that resulted in a decision to recommend or reject the measure. A decision to reject or accept each measure was made based on unanimous agreement among the investigators and consultants.

**Phase Two: Repeat Search and Review.** Following the acceptance of each measure, another structured literature search was performed. A MEDLINE search using the Ovid search engine was performed of English-language articles from 1959 to 1996 inclusive using the title of the measure as a search term and searching in the title, abstract, and body of an article to find all manuscripts containing that measure. A single investigator reviewed the title list generated from this search, and titles dealing with the development or evaluation of the measure were selected. A single investigator reviewed abstracts of these articles. Finally, a careful review of all article references was performed, appropriate additional references were identified, and the full-length articles were reviewed. Results of this review were then discussed with the other investigators and a consensus reached on measures to recommend. Findings from all newly found articles were also discussed. No new measures were identified. These findings were presented to the project consultants. The consultants unanimously supported the recommendations and had no suggestions for additional measures.

**Phase Three: Repeat Search and Review.** A second structured literature search was performed in October 2003 by the investigators (SMK, DWS, RFM). A MEDLINE search using the Ovid search engine was performed of articles from 1997 to October 2003 inclusive combining the original MeSH search terms with the RAMs and OMs from the initial search as shown in Figure 1. This search yielded a set of references that were limited to English language, human subjects, and publication in AIM journals. This created a reference set of 215 references. The titles were reviewed in the same structured manner as the initial review. Eleven titles were selected for further review. Abstracts of the selected references were reviewed. Reports not evaluating the feasibility, reliability, or validity of RAMs or OMs were excluded. Three abstracts were selected for full-article review. Examination of these articles with the same structured review conducted in Phase One resulted in the rejection of all three RAMs after group discussion.

## RESULTS

The RAMs and OMs identified are listed in Table 1. Of those identified and evaluated, only four will be recommended for use in out-of-hospital research and care of patients with respiratory distress. Those rejected failed, under structured review and consensus panel discussion, to be measures to be recommended for out-of-hospital use.

**Peak Expiratory Flow Rate.** Peak expiratory flow rate (PEFR) is one of a group of forced expiratory flow measures taken at the point of total lung capacity or at the point of maximal inspiration.\(^{15,16}\) Wright and McKerrow first described the measurement of the PEFR and the instrument used to obtain the measurement in 1959.\(^{17}\) The PEFR is primarily an index of obstruction in large airways. Mechanisms affecting the PEFR are the various causes of airway obstruction such as bronchomotor tone, tracheal obstruction, impaired mucus clearance, and modest effects of altitude.\(^{18-23}\) The use of the PEFR to monitor both in-hospital and out-of-hospital patients with asthma has been widely studied.\(^{19,24-26}\) Its use in the treatment and evaluation of COPD has also been recommended.\(^{24}\) Studies in patients with CHF have found only mild decreases in PEFR in stable patients but higher decreases in patients with severe symptoms of CHF.\(^{27}\)

Since the first article by Wright and McKerrow, various other instruments have been developed to measure the PEFR. These have been inexpensive portable devices that have been calibrated using the Wright flow meter as a standard.\(^{28-35}\) A device that is frequently used today in the United States is the mini-Wright peak flow meter.\(^{28-31,34-36}\) Children as young as 4 years of age can be taught to use portable devices to measure PEFR.\(^{37,38}\)

The validity of the PEFR has been measured using the forced expiratory volume in 1 second (FEV\(_1\)) as a criterion standard. In general, the PEFR correlates well with FEV\(_1\).\(^{39-43}\) Di Maria et al. showed correlation between the PEFR and FEV\(_1\) ranging from 0.84 to

---

**TABLE 1. Outcome of Measures Reviewed**

<table>
<thead>
<tr>
<th>Measure Identified and Rejected</th>
<th>Measures Accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fischl Index (for asthma severity)</td>
<td>Pulse oximetry</td>
</tr>
<tr>
<td>Fischl Index: Weiss modification</td>
<td>Peak expiratory flow rate</td>
</tr>
<tr>
<td>Verbal Breathing Score</td>
<td>Visual analog dyspnea scale</td>
</tr>
<tr>
<td>Symptom Score: Asthma (Chan-Yeung)</td>
<td>Mortality</td>
</tr>
<tr>
<td>Symptom Score: Asthma (Apter)</td>
<td></td>
</tr>
<tr>
<td>Asthma Hospitalization Predictive Index</td>
<td></td>
</tr>
<tr>
<td>Asthma Severity Scale: 6 months to 17 years</td>
<td></td>
</tr>
<tr>
<td>Two-item Bedside Index (Rodrigo)</td>
<td></td>
</tr>
<tr>
<td>Multiatribute Asthma Symptom Utility Index</td>
<td></td>
</tr>
<tr>
<td>Integrated Therapeutics Group Asthma Short Form</td>
<td></td>
</tr>
<tr>
<td>Rhinoconjunctivitis and Asthma Symptom Score</td>
<td></td>
</tr>
</tbody>
</table>
0.93. Studies conducted in emergency departments (EDs) have found correlations ranging from 0.77 to 0.83 in asthma and 0.69 in COPD. Because of these studies, the authors believe that the PEFR has been shown to be feasible for the out-of-hospital setting.

**Pulse Oximetry.** Oxygen saturation represents the percent saturation of available bound hemoglobin. The pulse oximeter (SpO2) is a common noninvasive device that represents a combination of oximetry and plethysmographic technologies and measures oxygen saturation.

Comparisons of pulse oximetry with arterial blood gases have shown high accuracy and agreement in both adults and children. Commonly used pulse oximeters are accurate within ±2% (SD) from 70% to 100% saturation and ±3% (SD) from 50% to 70% saturation, with no specified accuracy below 50% saturation. Evaluations of the accuracy and feasibility of pulse oximetry in the ED and out-of-hospital setting have also been conducted. These have found the devices to be accurate and their use very feasible. Because of this broad body of literature, the authors believe that pulse oximetry has been shown to be feasible for the out-of-hospital setting.

Several potential problems with the use of pulse oximetry have implications for out-of-hospital use. The presence of carboxyhemoglobin or methemoglobin may result in inaccuracies. Ambient light, dark skin, nail polish, and motion artifact can all affect the accuracy of the measurement. Severe anemia may affect the accuracy of the measurement if the oxygen saturation is <75%. Furthermore, pulse oximetry requires a pulsating vascular bed. Therefore, it may be inaccurate in low flow states.

**Visual Analog Dyspnea Scale.** Dyspnea can be defined as the unpleasant sensation of labored or difficult breathing and is synonymous with the term “shortness of breath.” The visual analog dyspnea scale (VADS) is a self-reported measure of shortness of breath. The patient is asked to indicate the degree of shortness of breath experienced by marking the line at the level indicating his or her level of discomfort. One end represents “not at all breathless” and the other end “worst possible breathlessness.” Aitken was the first to show that a VADS was a valid measure of dyspnea using a horizontally oriented scale. Gift and colleagues improved on Aitken’s work by reorienting the scale to a vertical line and testing the scale in clinical situations using patients with asthma and patients with COPD. This reorientation facilitated ease and understanding of use by patients. Investigations have shown that the VADS is valid for children with fully developed communication skills.

The VADS has been validated in healthy volunteers, patients with asthma, patients with COPD, and patients with cancer, with a variety of underlying causes of dyspnea. Correlation between the VADS and PEFR is good, ranging from −0.72 to −0.85. The test–retest reliability of the VADS has been shown to be high. Similar to potential challenges with the out-of-hospital use of the visual analog scale for pain, the severity of a patient’s dyspnea and interventions in the out-of-hospital setting may prevent optimal use of the VADS. The VADS should be studied in the out-of-hospital setting to ensure its feasibility for use, but the investigators recommend it nonetheless as a proposed measure.

**Mortality.** Unlike all other measures discussed thus far, which can function as either a RAM or an OM, mortality is the only recommended measure that is exclusively an OM. Although long-term mortality is significantly more important, short-term mortality (alive or dead on arrival at the hospital) is easily obtainable by out-of-hospital personnel and an appropriate OM for the out-of-hospital unit of service. Accessing mortality information after arrival at the hospital requires data linkage.

**Disability, Satisfaction, and Cost Outcomes.** No specific OMs proven to be applicable to the condition of respiratory distress in the out-of-hospital setting were found on review of the literature on disability (functional outcomes), patient satisfaction, or cost–effectiveness. Review of the references showed either disease-specific measures that had been developed to evaluate changes in chronic conditions over long periods of time or generic health status measures that had never been evaluated for assessing care rendered in the field. Because of this, we will not recommend any of these OMs for application in out-of-hospital research.

**DISCUSSION**

We recommend four measures that hold significant promise as priority RAMs and OMs for current use: PEFR, pulse oximetry, VADS, and mortality. The first three measures can be used as both RAMs and OMs. Core RAMs discussed in EMSOP III included age, gender, race, ethnicity, vital signs, level of consciousness, Glasgow Coma Scale score, standardized time intervals, and EMS provider impression. The core RAMs and OMs are recommended to be used in addition to these specific ones in out-of-hospital care evaluation and research.

An example of how a measure could be applied within the “episode of care” is shown in the out-of-hospital outcomes model (Figure 2). Although there have been some out-of-hospital clinical investigations using pulse oximetry, PEFR, and VADS, their reliability has not been studied in the out-of-hospital setting. Therefore, it will be necessary to substantiate that these measures are feasible and reliable when used to
evaluate care rendered in the field. Particular attention should be paid to the impact of patient age on feasibility. Age, for example, might alter a patient’s ability to use a peak flow meter.

It is notable and encouraging that recent investigations conducted as part of the Ontario Prehospital Advanced Life Support (OPALS) Study have linked intermediate and distal outcomes to advanced life support interventions performed in the field. These have included several RAMs and OMs related to disability (functional measures) and discomfort. These OMs have included the SF-36, the Cerebral Performance Category, the Health Utility Index Mark III, the Functional Independence Measure, and self-reported symptom relief. The SF-36 has become a widely utilized and adapted measurement tool. It is a 36-item survey that was constructed to survey health status in the Medical Outcomes Study. The OPALS investigators published their methods of determining costs in their out-of-hospital research. They have reported cost-effectiveness in terms of dollars per life saved and dollars per quality-adjusted life year.

Because EMSOP evaluated the actual feasibility, reliability, and validity of the RAMs and OMs related to out-of-hospital care, the absence of prior studies evaluating intermediate and long-term outcomes prevent our recommending any of these measures. However, it is encouraging to see robust out-of-hospital controlled clinical trials utilizing these measures. Undoubtedly these studies will add much to our knowledge of the feasibility of their use in future EMS outcomes research.

Capnometry and capnography are emerging technologies that may become valuable RAMs and OMs in the out-of-hospital setting. Although their use is increasing in intubated and nonintubated patients, inadequate research was found to conclude that they are valid and reliable in the out-of-hospital setting for this purpose.

This report underscores a number of important research issues that need to be addressed: 1) the feasibility and reliability of the PEFR and VADS in the out-of-hospital setting and how these vary by provider level; 2) the limits of the VADS in the pediatric population; 3) the absence of any identified OM for disability, satisfaction, or cost; and 4) a lack of established association between these measures and distal outcomes. For example, it is unknown whether improvement in the oximetry values measured after out-of-hospital interventions is related to outcomes such as hospital admission rates, length of stay, functional disability, and return to work.

LIMITATIONS

This study has several limitations. Only English-language articles and AIM-listed journals were used in the MEDLINE search. Thus, we cannot be absolutely sure that we have identified all potentially meaningful measures. We believed that the strategy, which included full article and reference list review, was the most effective to locate the influential publications for our structured review. In addition, our methodology depends on the validity of the use of an expert panel. Such methodologies always raise the possibility of bias or other attributes that can compromise the validity of the recommendations.

CONCLUSIONS

We reviewed the literature pertaining to RAMs and OMs relevant to respiratory distress. Using a previously published process, a structured review of studies, and consensus expert panel opinion, we recommend utilization of the following measures, along with core measures, as RAMs and OMs: pulse oximetry, PEFR, and VADS. Furthermore, we recommend mortality as an OM. The paucity of reliable, validated RAMs and OMs for outcomes research in the field, while not surprising, is disturbing. Major efforts are needed in the future to promote the development of validated tools for out-of-hospital outcomes research.

References