ORIGINAL CONTRIBUTIONS

LONG-TERM AIR MEDICAL SERVICES SYSTEM PERFORMANCE USING APACHE-II AND MORTALITY BENCHMARKING

Robert Silbergleit, MD, Richard E. Burney, MD, Kris Nelson, RN, Michael Snyder, RN, Mark J. Lowell, MD, Ronald F. Maio, DO

ABSTRACT

Objective. Air medical transport programs have been in existence for two decades. During this time, no outcome measures have been developed for these services. The authors examined severity scoring and mortality data from their air medical service to characterize its performance and to identify trends in acuity and mortality over a 15-year period. **Methods**. APACHE-II scores derived at the time of transport and hospital mortality data have been concurrently recorded in the flight database for adult transports since 1986. The authors analyzed these data and examined the correlation between APACHE-II score at the time of transport and hospital mortality for the 15-year period 1986–2001. **Results**. 13,808 adult transports were identified. APACHE data were available for 8,204 patients (59%) and mortality for 10,845 (79%), respectively. The number of transports increased from 935 to 1,231 per year. Mean APACHE-II for all patients was 11.6 ± 8.4. Overall mortality was 22%. Both patient acuity and mortality were trending upward over time. The correlation between APACHE-II and mortality was close and linear (mortality = 0.018 • APACHE-II – 0.0243, $R^2 = 0.97$). **Conclusions**. Both severity of illness and mortality of air-transported patients appear to be increasing slowly over time in response to changes in the health care system. The strong correlation between APACHE-II performed at the time of transport and mortality validates this technique for benchmarking. The slope of this correlation is an outcome-based characteristic of system performance that may allow monitoring of a system over time and comparisons between systems. **Key words**: air ambulance; APACHE; hospital mortality; transportation of patients.

PREHOSPITAL EMERGENCY CARE 2003;7:195–198

Although there has been increasing recognition of the importance of outcome measures in the study of health care delivery systems, the care provided by air medical transport services has not been consistently evaluated either quantitatively or qualitatively. A simple first step toward this end is measurement of the severity of illness or injury among these patients, and examination of the relationship between severity and mortality in the patients wgi have been transported.

Severity scoring systems assign points derived by regression analysis for abnormalities in physiologic measures, laboratory data, and patient history, and are used to predict mortality in patient populations. The Acute Physiology and Chronic Health Evaluation (APACHE) system for illness severity measurement was developed by Knaus et al. in 1981[1] and revised in 1985.[2] Physiologic severity scoring in patient transport using variants of the APACHE methodology was advocated and described by Bionet al. in 1985,[3] and then in the United States by Rhee et al.[4,5]

Despite these early efforts, little has been written about characterizing severity of illness in air medical patient transport over the past decade and a half. We have kept consistent acuity statistics on all adult transports since 1986. We undertook the present study to address the following questions: 1) Has the acuity of transported patients changed over time? 2) Has the mortality rate of transported patients changed over time? 3) Is there a relationship between APACHE-II score and mortality in transported patients? 4) Can acuity and mortality measures be used to benchmark system performance? The overall goal was to examine the use of physiologic severity scoring and mortality data to identify trends in, and characterize the long-term performance of, our air medical system.

Received July 19, 2002, from University of Michigan Departments of Emergency Medicine (RS, MJL, RFM) and Surgery (REB), and University of Michigan Survival Flight (KN, MS, MJL), Ann Arbor, Michigan. Revision received November 22, 2002; accepted for publication November 25, 2002.

Presented as a poster at the National Association of EMS Physicians annual meeting, Tucson, Arizona, January 2002.

Address correspondence and reprint requests to: Robert Silbergleit, MD, Department of Emergency Medicine, UMMC TC B1354/0303, 1500 East Medical Center Drive, Ann Arbor, MI 48109-0303. e-mail: <robie@umich.edu>.

METHODS

Survival Flight is the air medical and critical care transport service of the University of Michigan. In 1983 when the program began, it operated with one helicopter and did not have capability for balloon pump or extracorporeal membrane oxygenation (ECMO) transports. It currently operates three Bell 430 helicopters, a fixed-wing aircraft, and a ground ambulance, and performs both interfacility and scene transports. Specialty capabilities of the service include neonatal care, balloon pump, and adult and pediatric ECMO transports.

APACHE-II scoring of all adult transports by University of Michigan Survival Flight was initiated in 1986, the third year of program operation. It was recognized at that time that a consistent measure of the severity of illness of transported patients was needed for quality, research, and management purposes.

All flight nurses were trained in APACHE scoring, and forms were filled out at the completion of each adult transport, along with other patient documentation, for entry into the flight database. Flight nurses recorded APACHE-II scores immediately after transport. Intensive care unit (ICU) and hospital length of stay (LOS) and mortality data for patient transports were also routinely obtained in follow-up and entered into the flight database.

We queried this database for all adult air transports during the 15-year period 1986 through 2001. All ground and fixed-wing transports were excluded. Because severity scoring was not performed on pediatric transports, this patient group was excluded. All patient identifiers were removed from the data to be analyzed. The University of Michigan Medical Institutional Review Board determined that this analysis was exempt from review.

Each measure was evaluated for compliance in recording. For each calendar year, descriptive statistics for APACHE-II included the mean, median, and first and third quartile values. Patients with an expected mortality of greater than 10 percent (APACHE-II > 10[6]) were classified as critically ill. Annual mortality and proportion of critically ill are described as percentages of patients for whom data were available. Data was analyzed for trends and for correlation between APACHE-II score and mortality.

The correlation between APACHE-II scores calculated at transport and subsequent mortality was examined to determine whether the use of APACHE-II scoring in this context is valid, and to characterize the performance of the system. Severity scores from the entire data set were collapsed into strata of five points (1 to 5, 6 to 10, 11 to 15, etc.) and were plotted against the number of patients within the stratum, and the percent mortality for the stratum. Linear regression of mortality and APACHE-II scores was performed. Receiver operating characteristic (ROC) curves were plotted and the area under the curve was measured as a secondary description of the APACHE-II validity to allow comparison with prior published data.

RESULTS

We identified 13,808 adult transports between 1986 and 2001 from the database. APACHE-II scores and mortality data were available for 8,204 (59%) and 10,845 (79%) of patients, respectively. Survival Flight transported 935 patients in 1986; by 2000, this number had increased to 1,231 patients, an increase of 24%. In that 15-year period, the average APACHE-II score of transported patients was 11.6 ± 8.4, with a range of 0 to 54. The median was 9; the first and third quartiles were 5 and 17, respectively. Overall inhospital mortality was 22%. The average ICU LOS was 5.6 days; the average total hospital LOS was 12.7 days. Consistent with national trends, between 1986 and 2000, the ICU LOS decreased from 6 to 4.5 days and the hospital LOS decreased from 16.4 days to 9.3 days.

Severity score criteria for critical illness (APACHE-II > 10) were met by 48% of the patients. Both patient acuity and mortality have fluctuated over time, but both have trended upward, as shown in Figure 1.

The correlation between APACHE-II scores and mortality was both close and linear (mortality = 0.018 • APACHE-II score –0.0243, R^2 = 0.97) as shown in Figure 2. The resulting ROC curve is shown in Figure 3 and has an area under the curve of 0.875. The APACHE-II scores representing different points on the ROC curve are shown next to the curve.

DISCUSSION

One of the challenges that has been faced by air medical services since their inception in civilian transport has been finding a way to measure and characterize the severity of the diverse populations of critically ill or injured patients being transported. Early efforts to characterize patient severity used the Therapeutic Intervention Severity Score (TISS).[7] A 1988 report measuring TISS on 1,081 patients transported by five flight programs was the first attempt to compare program operations by patient acuity.[8] A subsequent study in 1990 evaluated 1,927 patients transported by six flight programs by multiple severity measures including APACHE-II.[9] Both of these studies found substantial variations among programs in both patient mix and severity that have never been further studied or explained, leaving unanswered questions about how to measure, compare, or monitor air medical system performance.

This report describes the largest existing data set of physiologic severity scoring by air medical flight teams. The chief findings of this study are that the acu-

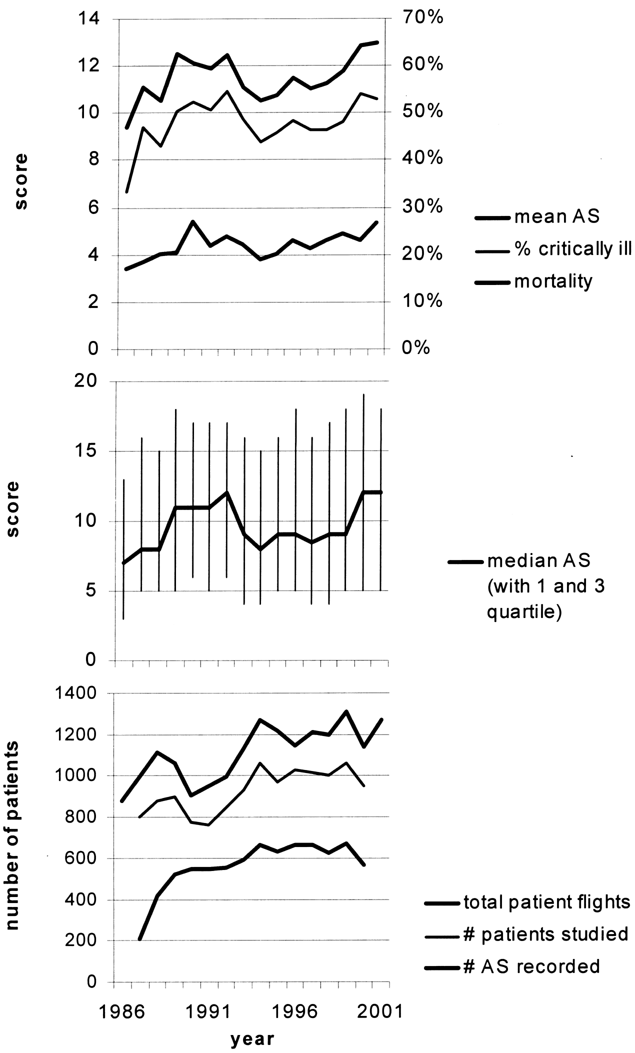

FIGURE 1. Mean APACHE-II score (AS, left axis), mortality (%, right axis), and proportion of critically ill defined as AS>10 (%, right axis) are shown at top; median AS with first and third quartile (vertical bars) are shown in the middle; and the number of patient flights, the number of patients studied, and the number of patients for whom an AS was recorded are shown at the bottom for the years shown.

ity of the air medical transport patients in our program has slowly increased over time, and that the mortality rate of transported patients has remained constant and proportionate to the increase in severity as measured by APACHE-II. The strong linear correlation between mortality and flight team APACHE-II scoring at the time of transport gives strong support to the hypothesis that this is a valid technique for monitoring system quality and performance and for establishing benchmarks. The relationship between APACHE-II and survival in these 8,204 patients is nearly identical to that previously described in 3,350 ICU patients.[6] In fact, the ROC curve defined by our experience shows that the predictive value of APACHE-II documented by the flight team in our system is identical to or slightly better than that described in 332 ICU patients scored by four specially trained physicians.[10]

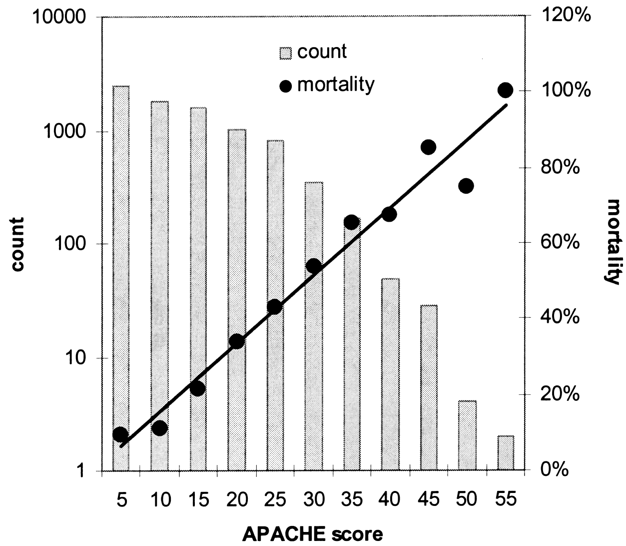

FIGURE 2. The number of patients stratified by APACHE score (*grey bars*) is shown along the left axis. Mortality (%, *black circles*) stratified by APACHE is shown with its linear regression (%, *black line*, R^2 = 0.97) along the right axis.

Another possible benchmark is the proportion of patients who could be considered "critical" by APACHE-II criteria. The proportion of critically ill transports, as measured by APACHE-II, has also slowly risen in our program. In one multisystem study, however, the mean APACHE-II score across six flight programs was 10.6,[9] compared with 11.6 in our program. Patients in that study were classified as critically ill if they had an estimated mortality of greater than 10%. Only 26% of the transported patients met this criterion by APACHE-II score. In our program, this number was 48%. Again, comparison data would be interesting for benchmarking purposes and could be another possible means for monitoring appropriateness of system performance.

Prospectively recorded severity or patient acuity data, such as these data from our program, may be useful for a variety of purposes. They demonstrate, at least for our program, a consistency of decision making regarding selection of patients for transport, and argue against the hypothesis that the increasing number of flight programs in the 1980s and 1990s has led to "overtriage" and inappropriate transports as a result of competitive pressures.[11] Comparable data from other programs would be valuable for comparison.

LIMITATIONS

There are limitations to this work that should be acknowledged. With APACHE-II data on 59% of patients and mortality data on 79%, our data collection was less than complete. We cannot be certain that some element of bias was not introduced by this incomplete data collection. Moreover, although the

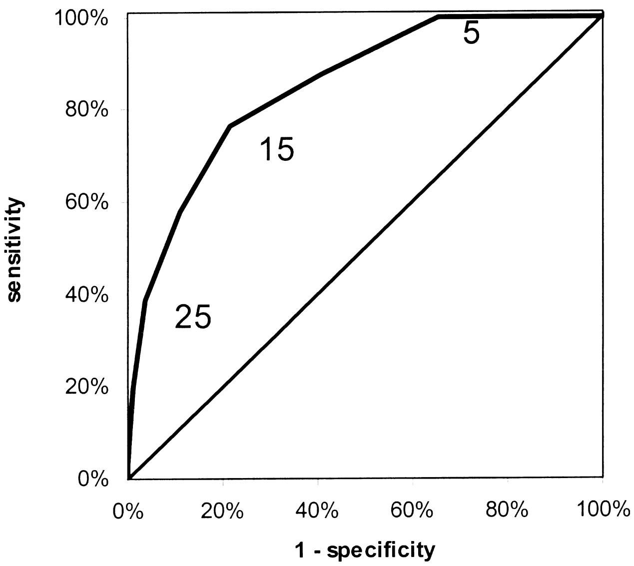

FIGURE 3. The receiver operating characteristic (ROC) curve for APACHE-II score prediction of mortality. The sensitivity and specificity for using APACHE-II score cutoffs of 5, 15, or 25 to predict mortality are shown with the associated cutoff. The area under the ROC curve is 0.875.

correlation between severity and mortality is clearly an outcome-based characteristic of system performance, the contribution of the flight program performance to overall health care system performance and patient mortality may be relatively small and confounded by other elements of care.

Despite these limitations, the clinical outcomes of patients transported by air medical services must be the fundamental measure of the value of air medical transport. These data suggest a method using physiologic severity scoring that provides a first step to studying the clinical outcomes of air medical transports in a meaningful, risk-stratified manner. We encourage other flight programs to adopt this methodology, so that performance can be studied across the air medical field.

CONCLUSIONS

Physiologic severity scoring by flight teams is a valid technique that characterizes the patient population and performance of a flight program over time. Despite increasing numbers of annual patient transports, transported patients are increasingly more severely ill by APACHE-II criteria. The proportion of patients deemed critically ill confirms this. Wider use and analysis of severity scoring will allow better comparisons of flight programs, and is a necessary first step to improved study of patient outcome measures in air medical transport.

References

1. Knaus W, Zimmerman J, Wagner D, Draper E, Lawrence D. APACHE—Acute Physiology and Chronic Health Evaluation: a physiologically based classification system. Crit Care Med. 1981; 9:591-7.
2. Knaus W, Draper E, Wagner D, Zimmerman J. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13: 818-29.
3. Bion JF, Edlin SA, Ramsay G, McCabe S, Ledingham IM. Validation of a prognostic score in critically ill patients undergoing transport. Br Med J (Clin Res Educ). 1985;291:432-4.
4. Rhee KJ, Fisher CJ Jr, Willitis NH. The Rapid Acute Physiology Score. Am J Emerg Med. 1987;5:278-82.
5. Rhee KJ, Mackenzie JR, Burney RE, et al. Rapid acute physiology scoring in transport systems. Crit Care Med. 1990;18:1119-23.
6. Rogers J, Fuller DH. Use of daily Acute Physiology and Chronic Health Evaluation (APACHE) II scores to predict individual patient survival rate. Crit Care Med. 1994;22:1402-5.
7. Cullen D, Civetta J, Briggs B. Therapeutic Intervention Scoring System: a method for quantitative comparison of patient care. Crit Care Med. 1974;2:57-60.
8. Burney RE, Rhee KJ, Cornell RG, Bowman M, Storer D, Moylan J. Evaluation of hospital-based aeromedical transport programs using therapeutic intervention scoring. Aviat Space Environ Med. 1988;59:563-6.
9. Rhee KJ, Baxt WG, Mackenzie JR, et al. Differences in air ambulance patient mix demonstrated by physiologic scoring. Ann Emerg Med. 1990;19:552-6.
10. Castella X, Gilabert J, Torner F, Torres C. Mortality prediction models in intensive care: Acute Physiology and Chronic Health Evaluation II and mortality prediction model compared. Crit Care Med. 1991;19:191-7.
11. Brathwaite CE, Rosko M, McDowell R, Gallagher J, Proenca J, Spott MA. A critical analysis of on-scene helicopter transport on survival in a statewide trauma system. J Trauma. 1998;45:140-4; discussion 144-6.