Journal of Critical Care (2009) **24**, 401–407

ELSEVIER

Journal of Critical Care

Reliability of intensive care unit admitting and comorbid diagnoses, race, elements of Acute Physiology and Chronic Health Evaluation II score, and predicted probability of mortality in an electronic intensive care unit database ☆

Joshua B. Wenner BA [a], Monica Norena MSc [a], Nadia Khan MD, MSc [a,b], Anita Palepu MD, MPH [a,b], Najib T. Ayas MD, MPH [b], Hubert Wong PhD [a,c], Peter M. Dodek MD, MHSc [a,b,*]

[a] *Center for Health Evaluation and Outcome Sciences, St. Paul's Hospital, Vancouver, B.C. Canada V6Z 1Y6*
[b] *Department of Medicine, University of British Columbia V5Z 1M9*
[c] *School of Population and Public Health University of British Columbia V6T 1Z3*

Keywords:
Severity of illness index;
Reproducibility of results;
Observer Variation

Abstract

Background: Although reliability of severity of illness and predicted probability of hospital mortality have been assessed, interrater reliability of the abstraction of primary and other intensive care unit (ICU) admitting diagnoses and underlying comorbidities has not been studied.

Methods: Patient data from one ICU were originally abstracted and entered into an electronic database by an ICU nurse. A research assistant reabstracted patient demographics, ICU admitting diagnoses and underlying comorbidities, and elements of Acute Physiology and Chronic Health Evaluation II (APACHE II) score from 100 random patients of 474 admitted during 2005 using an identical electronic database. Chamberlain's percent positive agreement was used to compare diagnoses and comorbidities between the 2 data abstractors. A κ statistic was calculated for demographic variables, Glasgow Coma Score, APACHE II chronic health points, and HIV status. Intraclass correlation was calculated for acute physiology points and predicted probability of hospital mortality.

Results: Percent positive agreement for ICU primary and other admitting diagnoses ranged from 0% (primary brain injury) to 71% (sepsis), and for underlying comorbidities, from 40% (coronary artery bypass graft) to 100% (HIV). Agreement as measured by κ statistic was strong for race (0.81) and age points (0.95), moderate for chronic health points (0.50) and HIV (0.66), and poor for Glasgow Coma Score (0.36). Intraclass correlation showed a moderate-high agreement for acute physiology points (0.88) and predicted probability of hospital mortality (0.71).

Conclusion: Reliability for ICU diagnoses and elements of the APACHE II score is related to the objectivity of primary data in the medical charts.

☆ Research done at Center for Health Evaluation and Outcome Sciences, St Paul's Hospital, Vancouver, BC, Canada.

* Corresponding author. Center for Health Evaluation and Outcome Sciences, Vancouver, British Columbia, Canada V6Z 1Y6.
E-mail address: pedodek@interchange.ubc.ca (P.M. Dodek).

0883-9441/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcrc.2009.03.008

1. Introduction

The Acute Physiology and Chronic Health Evaluation II (APACHE II) score is one of the most commonly used measures to quantify severity of illness and to predict hospital mortality for intensive care unit (ICU) patients [1]. Probability of mortality values derived from APACHE scoring is used as a quantifiable benchmark for determining and comparing ICU performance; therefore, the reliability of this measure is particularly important [2]. Efforts have been made to measure reliability of the APACHE score using a variety of techniques. Some researchers have done interrater reliability studies using actual patient data that have varied in size and scope, from small case-mix [3,4] to larger-scale analyses [5]. Others have measured interrater reliability among researchers who are provided with a set of fictitious patient information [6]. In general, most of the research on this subject has assessed the reliability of both APACHE II scores and the probability of hospital mortality but without any particular focus on ICU admitting diagnoses. Findings from these studies show variability among some components of the APACHE II score, which could subsequently affect confidence in calculations of predicted mortality.

Using the APACHE methodology, probability of mortality is derived from the APACHE II score and a coefficient that is based on the primary admitting diagnosis. Determining a single admitting diagnosis is a subjective task; different interpretations of what is the primary admitting diagnosis will lead to different coefficients and different predicted probabilities of mortality. Furthermore, patients often have more than one reason for ICU admission and they come to the ICU with a variety of important comorbid diagnoses. These diagnoses may be used to further stratify patients and to calculate comorbidity indices. Interrater reliability in the abstraction of other admitting diagnoses and underlying comorbidities along with APACHE II variables and patient demographics has not been assessed. Therefore, the aim of this study was to determine interrater reliability for primary admitting, other admitting, and comorbid diagnoses as well as for race, elements of the APACHE II score, and predicted probability of mortality for patients admitted to a medical-surgical ICU.

2. Methods

2.1. Setting and data sources

This study was conducted in a 15-bed medical-surgical ICU within a 350-bed tertiary teaching hospital in Vancouver, British Columbia. Baseline patient data, including elements of the APACHE II score, as well as diagnosis and procedure information for each ICU patient encounter, were originally abstracted from medical charts and entered into a Web-based electronic data base (Oracle Database 9i, Redwood Shores, Calif) from January 1 to December 31, 2005, by a trained ICU nurse-research coordinator. All diagnostic categories were based on the Intensive Care National Audit and Research Center dictionary of diagnoses [7].

2.2. Study design

To determine the reliability of APACHE II and diagnosis information, we compared existing baseline data in the electronic database to data collected from the same sources by a second data collector from July 1 to August 31, 2006. The second data collector was trained how to review charts, used identical definitions of data elements including diagnoses and procedures, and independently and systematically reabstracted medical chart data into an identical electronic database. Data collectors were blinded to each other's data entry, and no additional data were available at the time of reabstraction. Of 474 patients admitted to the ICU during the calendar year 2005, 100 records were randomly selected for reabstraction. This sample size was chosen to achieve an SE of less than 0.1 for the κ statistic based on the assumption that for a given comorbid condition (eg, HIV), each rater found a prevalence of 8%, disagreements were symmetric across the raters, and the observed raw proportion of agreement was 98% [8]. Consistent with the original definition of the APACHE II score, records from patients younger than 18 years (n = 0) or having an ICU stay less than 8 hours (n = 3) were excluded [1]. We also excluded records from the original abstraction that were incomplete (n = 2), leaving 95 of the original 100 records for reabstraction. From these 95 records, the second data collector conducted a structured review, recording patient name, sex, race, date of birth, date/time of ICU admission, primary and other ICU admitting diagnoses, underlying comorbidities, and all individual elements of the APACHE II score.

Items abstracted for calculation of the APACHE II score were from the first 24 hours of intensive care for each patient [1]; this period was defined according to the first recorded time of ICU admission. As pharmacologic sedation can obscure Glasgow Coma Score (GCS) assessments, if GCS or GCS components were not recorded in the medical chart, a GCS score of 15 was assigned to patients (consistent with practice of the original data collector). Both the original and the second rater assigned no APACHE II points for those elements of the score that were not available.

Primary and other admitting diagnoses, as well as underlying comorbidities, were reabstracted from progress notes, ICU admission orders, multidisciplinary flow sheets, and consultation records. The electronic database required only one primary admitting diagnosis to be entered to calculate a predicted probability of hospital mortality [1]. Guidelines were also necessary for determining and differentiating similar or potentially interchangeable diagnoses. Therefore, septic shock and septicemia, as well as all types of pneumonia, were included under the headings of sepsis and pneumonia, respectively.