**Purpose:** Retrospective cohort studies involving intensive care unit (ICU) patients frequently describe acuity of patient illness with severity classification scores. Most risk scores use a combination of physiologic variables and comorbidities to calculate the score. While physiologic variables can often be extracted from electronic medical records (EMR), identifying comorbidities has traditionally relied on manual review of clinical documentation. This process is time-consuming and tedious with larger sample sizes, thus an automated electronic scoring tool using diagnostic coding would facilitate the process. Our study objective is to validate the use of diagnosis codes in the calculation of severity scores.

97/100 words

**Methods:** Patients admitted to an ICU between 07/01/14 and 06/30/16 were screened for the study. Patients were excluded if they were pregnant, a prisoner, or were missing data necessary to calculate a risk score. Sixty patients were then randomly selected for inclusion. Physiological and comorbidity data was extracted and manually collected for all patients in this IRB-approved, retrospective, observational study. Three severity scores were calculated: the Acute Physiology and Chronic Health Evaluation (APACHE) II, the APACHE III, and the Simplified Acute Physiology Score II (SAPS II). Each score was calculated using eight different comorbidity definitions utilizing International Statistical Classification of Diseases and Related Health Problems(ICD) codes, which were evaluated for accuracy and reliability against manual review of clinical documentation. Comorbidity sets used included a set generated by the authors through validated literature and manual collection, Agency for Healthcare Research and Quality (AHRQ), Elixhauser Comorbidity Index, and Quan’s comorbidity set. Each data set included a version with and without the use of Diagnosis-Related Group (DRG) codes to exclude comorbidities related to the admitting diagnosis. The primary outcome was to determine the difference in score between electronic and manual methods amongst data sets. Secondary outcomes include difference in predicted mortality and proportion with different comorbidities from the manual set. Comparison of severity scores was performed using paired t-test and Wilcoxon signed rank test.

221/225 words

**Results:** There was no difference in the APACHE II or SAPS II severity scores for each of the eight comorbidity sets when compared with the manual set. A statistically significant difference was found in the APACHE III score for all eight comorbidity sets compared with the manual set (p-value equals 0.02 for each set). In all cases where the score differed among all severity scores, the score using diagnostic coding was higher than the score for the manual set. Among patients with a different score from manual, the median (interquartile range [IQR]) difference for all sets was 5 (2-5) for APACHE II, 11 (7-13.5) for APACHE III, and 9 (9-9.5) for SAPS II. Predicted hospital mortality calculated using the SAPS II regression model equation was not significantly different for each set compared with manual. The proportion of patients with different comorbidities identified from manual was highest using APACHE II criteria (28.3 to 31.7 percent among the eight comorbidity sets), followed by APACHE III (16.7 percent differed from manual for all sets), and SAPS II (5 percent differed for all sets). There was no intra-set difference in comorbidities identified using APACHE III or SAPS II criteria.

194/200 words

**Conclusion:** These results show that diagnostic coding can reliably be used to identify patient comorbidities for the calculation of severity classification scores. Despite some variation in identification of comorbidities, there was no difference in APACHE II and SAPS II scores. Although finding a significant difference in APACHE III score, this difference was driven by comorbidity variation in a small number of patients and may not be clinically significant. The use of an electronic scoring tool based on physiologic and diagnostic data extracted from the EMR can be an efficient way to calculate severity classification scores in larger retrospective studies.

97/100 words