Lessons from Evaluating an Automated Patient Severity Index

RICHARD F. GIBSON, MD, PHD, PETER J. HAUG, MD, SUSAN D. HORN, PHD

Abstract  Objective: To report lessons learned from evaluation of an automated interface between a hospital clinical information system and a severity of illness index.

Design: A system was developed to convert coded electronic patient findings from the HELP System at LDS Hospital into the attributes used by the Computerized Severity Index (CSI) to calculate a severity of illness score. Performance was assessed by comparing the automated CSI score with the manual CSI score (from paper chart review) and by evaluating changes introduced by augmenting the manual CSI score with verified patient data discovered by the automated CSI method.

Measurements: The strengths and weaknesses of each method are presented.

Results: The automated CSI score matched the manual CSI score in 61% of the cases. Sources of errors were analyzed. When the automated score was in error, two-thirds of the time it was due to the lack of codes in the HELP system representing CSI concepts; one-third of the time it was due to nurses not using established HELP system codes. Surprisingly, significant problems were also discovered in the manual system, making it difficult to define a "gold standard."

Conclusions: Automated computerized severity indices have great potential for future applicability once their performance exceeds that of the time-consuming manual chart review method. Neither automated nor manual methods are adequate at the present time. This area remains a fertile ground for future research.


Severity indexes for hospitalized patients have been studied for 20 years. They have been used in examining mortality, \textsuperscript{5,6} readmission and reoperation rates, \textsuperscript{5} length of hospital stay, \textsuperscript{3,4,6–9} and hospitalization costs. \textsuperscript{5,6,9–12} They have also been used in adjusting for patient severity in clinical studies. \textsuperscript{13–15} Some severity

Affiliations of the authors: Division of Information Services, Providence Health System, Portland, OR (RFC); Department of Medical Informatics, LDS Hospital, University of Utah, Salt Lake City, UT (PJH, SDH).

This work was supported in part by National Library of Medicine Grant Number F37 LM00012, awarded by the National Institutes of Health, Department of Health and Human Services, and the Department of Medical Informatics, LDS Hospital, University of Utah.

Correspondence and reprints: Peter J. Haug, MD, Department of Medical Informatics, LDS Hospital, Eighth Avenue and C Street, Salt Lake City, UT 84143.

e-mail: ldphaug@ihc.com

Received for publication: 9/7/95; accepted for publication: 5/20/96.

of illness indices have been based on the patient's age, sex, discharge diagnoses, and procedures performed; this information has been termed "discharge abstract" data because the data are found on the front page of most hospital paper records. These data are collected in computerized form uniformly in essentially all hospitals in this country. Other severity indices use clinical patient descriptors in addition to discharge abstract data. These clinical severity systems can group patients homogeneously for the above-noted purposes, but clinical data collection from review of paper charts has been tedious and expensive. Severity of illness indexes based on clinical patient data that were easily and inexpensively obtained would be useful in patient care and health services research.

The Health Evaluation through Logical Processing (HELP) system at LDS Hospital \textsuperscript{16,12} has collected computerized clinical patient data for nearly 30 years. This paper describes the development and testing of an electronic interface between the HELP system and the Computerized Severity Index (CSI), a commercially
Background

There exists no "gold standard" for severity systems. Higher severity suggests a greater "disease burden" and more complicated management, both diagnostic and therapeutic. Although severity is a medical concept, severity systems have been developed for and validated by their ability to predict cost and length of stay. A severity system that is optimized to predict cost and length of stay will behave differently from one optimized to predict mortality. Therefore, a group of patients defined by a severity system may be homogeneous with respect to a given outcome but heterogeneous with respect to resource intensity.

Three severity indexes—Disease Staging, Patient Management Categories, and Acuity Index Method—all use discharge abstract data routinely collected by every hospital on the front sheet of the chart. At most hospitals, these data have already been captured in coded, electronic form readily accessible by a computerized version of a severity index. Three other severity indexes—MedisGroups, APACHE III, and CSI—all use clinical data about the patient, which can only be obtained by a review of the more detailed paper medical record. Manual chart review is more time consuming and expensive than using already-computerized discharge abstract data. Time estimates for chart abstraction for CSI and MedisGroups have been in the range of 20-30 minutes per chart. Completing such reviews would require one full-time equivalent for every 4,000-6,000 hospital discharges.

Available from International Severity Information Systems (ISIS), Inc., Salt Lake City, Utah, CSI maps each of the 12,775 International Classification of Diseases, Ninth Revision, Clinical Modification (ICD)-9-CM diagnoses to one of 833 matrices. Each matrix comprises 4 to 50 indicators; physiological patient attributes such as vital signs, physical examination findings, laboratory values, and diagnostic studies. Each indicator is broken down into "bins," to which are assigned the severity levels 0 through 4. Level 0 means that there is no evidence to support the indicator; Level 1 indicates normal or mild symptoms; Level 2, moderate illness; Level 3, severe signs or symptoms; and Level 4, catastrophic or life-threatening signs or symptoms.

The HELP system at LDS Hospital, a 520-bed tertiary care center in Salt Lake City, has been under development since 1967. It provides an integrated, computerized environment for using and developing clinical, administrative, and financial modules. An integrated expert-system tool supports medical decision making. The HELP system has collected clinical findings about patients using hierarchical Pointer-to-PTXT (PTXT) codes. The authors were intrigued by the possibility of performing CSI scoring automatically by correlating PTXT codes with the CSI bins.

The patient attributes on the HELP system are fairly atomic; they cannot be subdivided easily into multiple individual "sub" findings. Examples of atomic findings include the following: 1) a nurse may enter a patient finding such as "rales in right lung," or "Foley catheter urine output: 250 ml," or "purulent sputum" with the timestamp "17-Dec-94 1730h" into the patient database; 2) the laboratory information system may enter "hemoglobin: 12.1" with a timestamp; and 3) the automatic chest x-ray report interpreter may enter "chest x-ray infiltrate" with a timestamp.

Some of the CSI indicator bins correspond directly with PTXT codes. For example, the bins for lowest systolic blood pressure, highest temperature, and highest white blood cell count can all be satisfied by a single PTXT code for each bin. Many of the CSI bins represent such abstract concepts as (1) "frequent bloody diarrhea," where the CSI manual defines "frequent" as greater than five episodes in 24 hours; (2) "rales in three or more lobes"; (3) "vomiting after a meal"; and (4) "fever greater than 101° F, on two occasions greater than four hours apart, with negative bacterial cultures." The authors had to encode logic to group several atomic PTXT codes to represent the more abstract concepts of the CSI bins.

Methods

Mapping CSI to the HELP System

To map CSI to the HELP system, project members examined each CSI indicator, bin by bin, and mapped them to any corresponding PTXT codes in use in the HELP system. This required medical judgments as to whether a patient attribute identified by a CSI bin was addressed by any PTXT code(s). A project-developed "C" program culled all demographic data, discharge diagnoses, procedures, and PTXT codes judged pertinent to CSI scoring from the mainframe-based HELP patient database. The resulting text file contained only three elements for each patient for each clinical event: the PTXT code, the data value (e.g., the numeric value of the blood pressure or temperature), and the timestamp (date and time in minutes). The text file with patient data was then transferred over the local-area network to an IBM-compatible personal computer.
running Advanced Revelation, the database application environment used for CSI. An Advanced Revelation procedure opened the text file of patient data and processed one patient at a time.

The Advanced Revelation procedure implemented mapping logic to convert the findings on the HELP System to the concepts in CSI. All the CSI indicators having associated PTXT codes were represented with one of 13 different logic statements using Boolean operators and lists of PTXT codes pertaining to that indicator. Mapping composite CSI concepts required identifying several atomic PTXT codes in the patient database and using timestamps to verify that the codes referred to the same patient event. For example: the indicator “frequent bloody diarrhea” was represented by logic as follows: one PTXT code for “stool,” AND one PTXT code for “loose” OR “watery,” AND one PTXT code for “bloody,” AND all PTXT codes have the same timestamp, AND six or more such combinations within a 24-hour period. Scoring started with the indicator bin with the greatest severity level. If the associated PTXT codes could not be found, the program went to the next highest level bin and repeated the process. If no bin was satisfied, the indicator remained at Level 0.

Selection of the Experimental Sample

As of July 1994, LDS Hospital had obtained manual CSI scores on 3,372 patients hospitalized between January 1990 and September 1993. Those scores were based on a retrospective chart review conducted and manually CSI scored between May 1993 and July 1994. Patients on the obstetric, psychiatric, and rehabilitation floors were excluded from study because the HELP system on those nursing units was not used for nurse charting. Within each Major Diagnostic Category (MDC) of the ICD-9-CM hierarchy, authors randomly chose approximately 30 cases at each of the four manual CSI levels (there were no CSI Level 0 patients). If fewer than 30 cases were available at a given CSI level, all cases were selected for automated scoring. It was felt that only seven MDCs had enough cases across all four severity levels to justify inclusion in the study. The Endocrine and Urinary System MDCs had only 58 and 76 cases, respectively, but they were included to increase the breadth of the diagnoses and the number of cases under study.

The observed agreement between automated and manual CSI scores was tallied for the Admission period, the Maximum period, and the Discharge period and was then classified by each MDC. Agreement was analyzed by the weighted Kappa statistic.\textsuperscript{16,29}

Detailed Analysis of the Causes of Manual and Automated Disagreement

The authors attempted to identify factors limiting the performance of the automated interface to determine implications for automatic severity scoring for hospital information systems in general and for the HELP system in particular. The authors believed the best way to investigate these issues was to compare the electronic medical record (the source of data for automated CSI) with the paper medical record (the source of data for manual CSI). The authors noted all cases in which the automated score differed from the manual score for the Maximum period. Authors then randomly selected more than half of the cases in which the automated Maximum score differed from the manual score. For those cases, the authors obtained the full paper medical record from the Medical Records department and identified every individual indicator to determine the reason for the discrepancy in scoring. Figure 1 details the paper chart review process. Each indicator for which the automated level was different from the manual level was counted as one “discrepancy.”

Results

Evaluation of CSI Scoring in 716 Cases

Agreement was evaluated for 716 cases scored by both manual and automatic CSI methods for the Maximum period. The automated Maximum CSI score, when compared with manual CSI scoring, was lower in 30% of cases, the same in 61% of cases, and higher in 9% of the 716 cases. Automated–manual agreement was best in the levels at both ends of the range, Level 1 and Level 4. This result was expected because at either extreme the automated scores can differ in only one direction from the manual scores, and because of systematic biases inherent in the study. Agreement was best in the 185 Level 1 cases, as expected (automated scores lower in 0% of cases), the same in 90%, and higher than manual scores in 10%). Patients default to Level 1 when only normal CSI-related findings are present in the medical record. Level 4, with 196 cases, showed the next greatest agreement (automated scores lower in 33% of cases, the same in 67%, and higher than manual scores in 0%). This level of agreement was in part due to a systematic bias, which improved automated–manual agreement in Level 4 cases. Level 4 is assigned by CSI to any patient with in-hospital death. Automated CSI detected in-hospital death for 53 (27%) of the Level 4 patients through the discharge status obtained from the HELP system. When deaths were excluded, automated–manual agreement in Level 4 cases changed; 45% of cases had...
215 cases had
Maximum Auto CSI lower than
Maximum Manual CSI.

99 of the 215 cases were randomly
chosen for paper chart review of the
Maximum CSI score.

57 of the 99 cases had
Admission Auto CSI
lower than Admission
Manual CSI. All 57
cases had the paper chart
reviewed for the
Admission CSI score.

63 cases had
Maximum Auto CSI higher than
Maximum Manual CSI.

47 of the 63 cases were randomly
chosen for paper chart review of the
of the Maximum CSI score.

12 of the 47 cases had
Admission Auto CSI
lower than Admission
Manual CSI. All 12
cases had the paper chart
reviewed for the
Admission CSI score.

35 of the 99 cases had
Discharge Auto CSI
lower than Discharge
Manual CSI. All 35
cases had the paper chart
reviewed for the
Discharge CSI score.

64 of the 99 cases had
Discharge Auto CSI
equal to Discharge
Manual CSI. No paper
chart review was done.

24 of the 47 cases had
Admission Auto CSI
equal to Admission
Manual CSI. No paper
chart review was done.

None of the 99 cases had
Admission Auto CSI
higher than Admission
Manual CSI. No paper
chart review was done.

None of the 99 cases had
Discharge Auto CSI
higher than Discharge
Manual CSI. No paper
chart review was done.

57 charts reviewed in this
column for Admission CSI.

35 charts reviewed in this
column for Discharge CSI.

In summary: 146 different hospitalizations
146 different patients
146 Maximum period paper chart reviews
80 Admission period paper chart reviews
46 Discharge period paper chart reviews.

For the Admission period, the automated scoring produced the same score as the manual scoring in 388 (57%) of the 685 cases scored by both methods.30 (Not all of the 716 cases scored for the Maximum period were scored for the Admission and Discharge periods.) The automated score was lower than the manual score in 254 (37%) of the cases, and higher than the manual score in 43 (6%) of the 685 cases. In the Discharge period, the overall agreement rose to 503 (74%) of 680 cases.30 The automated score was lower than the manual score in 150 (22%) of the cases, and it was higher than the manual score in 27 (4%) of the 680 cases.

Possible Explanations for Differences in Manual and Automated CSI Scoring Results

Although the distribution of manual scores was relatively even (the authors picked approximately 30
cases at each CSI level for each MDC), the distribution of automatic scores was skewed to the lower severity levels because the HELP system did not have PTXT codes to represent all the abnormal patient findings in the paper chart. In addition, the overall agreement for the Admission (57%) period was less than that for the Maximum period because the 24-hour length of the former period decreases the chances that one of the nurses will observe and electronically chart an abnormal finding.

One abnormal finding at any time during the hospital stay can be used to score Maximum CSI even if that finding was not noted by any other clinician. Also contributing to lower Admission period agreement is the fact that an abnormal finding during that period is more likely to arise in the emergency department than would an abnormal finding over the entire hospital stay.

When the study patients were in LDS Hospital, the emergency department had no computerized nurse charting. Although the Discharge period was also only 24 hours long, Discharge score agreement was increased by the large number of Level 1 cases (427 of 680, 63%), which had 94% agreement. The Discharge period tends to have lower CSI score levels than any other period because patients tend to return to their usual states of health just before discharge.

Table 1

<table>
<thead>
<tr>
<th>Category of Discrepancy</th>
<th>Admission (80 cases)</th>
<th>Maximum (146 cases)</th>
<th>Discharge (46 cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Auto lower than manual</td>
<td>Auto higher than manual</td>
<td>Auto lower than manual</td>
</tr>
<tr>
<td></td>
<td>(69 cases)</td>
<td>(11 cases)</td>
<td>(99 cases)</td>
</tr>
<tr>
<td>Manual coder error</td>
<td>100</td>
<td>23</td>
<td>113</td>
</tr>
<tr>
<td>No PTXT defined</td>
<td>139</td>
<td></td>
<td>192</td>
</tr>
<tr>
<td>PTXT not used (source below)</td>
<td>67</td>
<td></td>
<td>87</td>
</tr>
<tr>
<td>MD dictated note</td>
<td>32</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>MD handwritten note</td>
<td>8</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>RN handwritten note</td>
<td>25</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>RN typed free text into HELP</td>
<td>2</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>327</td>
<td>28</td>
<td>436</td>
</tr>
</tbody>
</table>

Results of Detailed Analysis of Automated and Manual CSI Disagreement

The authors identified factors that limited the performance of the automated and manual approaches to CSI scoring (Table 1). Each indicator for which the automated level differed from the manual level was counted as one “discrepancy.” Each case with discrepancies had between one and eleven discrepancies. The authors grouped errors into five major etiologic categories, listed in Table 1 in descending order of frequency.

Manual Coder Error

Of the 452 discrepancies due to errors made by the manual CSI chart abstractors, 182 (40%) resulted from the reviewers overlooking a finding discovered electronically by the automated method. This type of error led to the automated CSI score being higher than the manual score, although the former was correct. The remaining manual coder errors all led to the manual score being higher than the automated score.

The manual reviewers misinterpreted findings clinically in 109 (24%) of the 452 errors (e.g., they noted respiratory distress when the patient simply had minor respiratory symptoms or rhonchi). Incorrectly calculating the temporal boundaries of the Admission, Discharge, and Postoperative periods accounted for...
The authors found no evidence in the paper chart for 65 (14%) of the errors. In 46 (10%) of the discrepancies, the coder interpreted a symptom reported by the patient as a finding. The scoring rules for CSI explicitly state that a clinician needs to observe the finding; patient reporting is not accepted for CSI scoring. In 18 (4%) of the errors, the manual coder misread the printed vital signs. Nine (2%) of the errors were numbers incorrectly typed into the CSI program. Errors were reviewed with the lead CSI abstractor at LDS Hospital so that the authors verified their interpretation of the error and so that the abstractors could get feedback on their performance.

No PTXT Defined

The greatest reason for the automated score being lower than the manual score was that CSI indicators lacked representation in the HELP system. Undefined PTXT was found to be the cause in 402 (44%) of 918 discrepancies (the total number of discrepancies contributing to the automated score being lower than the manual score). Manual CSI abstractors found evidence to support these indicators in paper documents derived from nurses, physicians, and other health professionals.

Existing PTXT Not Used

PTXT codes had been defined and were in use on HELP system nurse charting menus for concepts in this category. Failure to use the existing PTXT codes was the cause in 170 (19%) of 918 discrepancies leading to the automated score being lower than the manual score. The items were not selected by nurses because (a) their division did not have access to the HELP system (e.g., the emergency department), (b) the nurses noted the finding by handwriting a note or typing it into the HELP system as free text, or (c) the nurses did not note the finding, either because they chose not to or because the patient did not exhibit the finding. The authors did not include patients who spent a significant part of their hospital stay in nursing units without HELP system nurse charting (obstetrics, psychiatry, and rehabilitation). However, many of the patients studied entered the hospital via the emergency department, where there was no HELP system nurse charting. As expected, many patients exhibited their most severe abnormalities in the emergency department. Excluding patients who spent time in the emergency department would have significantly decreased the number of patients available for this study. Including such patients resulted in a systemic bias against automated CSI scoring.

The manual coders found evidence for more than half of the unused PTXT concepts in physicians' dictated and handwritten notes. Few of the currently used PTXT codes in the HELP system derive from physician-entered findings. Most of the remaining evidence was found in notes handwritten by the nurses. Nurses at LDS Hospital can store free text findings on the HELP system, but these findings are not linked to PTXT codes. Nurses are encouraged to use the nurse charting menu items (with their associated PTXT codes), but they are free to type free text notes into the HELP system or write notes in the paper chart. The nurses infrequently typed significant findings into the HELP system; either they used the menus with coded choices, or they handwrote the chart.

LDS Hospital and HELP System Errors

Seven percent (75 of 1,120) of all discrepancies were due to errors in hospital procedure unrelated to CSI scoring. During 1992–1993, a new type of electrocardiogram (ECG) machine was used at LDS Hospital, but the ECG machine–HELP interface was not updated until late 1993. Findings that would ordinarily have been captured by the electronic ECG interface were absent from automated CSI scoring for patients hospitalized during this period, but the findings were still printed in the paper chart.

Automated CSI found data in the electronic medical record that had not been printed for the paper chart. The nurses are responsible for printing their coded assessments; the Medical Records department is responsible for printing the final version of the laboratory studies and checking to see that nursing notes and radiology reports are present and complete before the chart is filed. It was not always possible for a chart abstractor to determine if pages of vital signs or laboratory studies were missing. Data on those pages were detected electronically but were missed manually.

Intensive Care Unit (ICU) vital signs are not printed as numeric characters; abstractors must estimate the values from relatively coarse, printed graphs. Such estimates are rounded from their true electronically stored values by as much as 7 mm Hg for blood pressure, 7 beats/minute for pulse rates, and 0.4°C for temperatures. Rounding errors in either direction led to scoring discrepancies.

Other Causes

Inadequacies in the automated CSI scoring method remained after optimizing the method to capture the greatest number of patient data with the fewest number of errors. Two thirds (14 of 21) of the miscellaneous...
ous discrepancies were due to unremedial deficiencies in the automated CSI method.

Figure 2 represents the causes of the automated score being lower than the manual score, as well as the corresponding percentage for each cause; Figure 3 represents the causes of the automated score being higher than the manual score (in 146 paper charts reviewed for the Maximum period) and the corresponding percentages.

A review of 146 of the 278 cases in which the automated scoring differed from the manual scoring identified 119 cases with correctable manual data. For each of these 119 cases, the authors manually entered the correct patient data into the previously manually scored case, replacing incorrect data entered by the abstractors. Rescoring the 119 cases created new, "revised" scores that agreed with the original automated method scores in 79% of the cases and with the original manual method scores in 80% of the cases.

**Discussion**

**No General Standard for Comparison**

The major—and somewhat surprising—finding of the present study is that there is no general standard against which to compare the performance of automated CSI systems. The hypothesis of the experiment was that the electronic patient database of the HELP system at LDS Hospital could be used to automatically derive a clinically based severity score. This study showed that automated and manual CSI methods both produced significant errors, enough that neither method could be considered a gold standard.

The authors were surprised by how many scoring discrepancies suggested that the manual scoring was incorrect for a given indicator. Until this discovery, the authors had felt that the major cause of differences between automated and manual scoring would be the absence of associated PTXT codes for CSI indicators, in which case the automated scoring would be incorrect. While only 146 of the 278 cases in which the automated scoring differed from the manual scoring were examined for possible correctable data, the manual scoring was determined to be incomplete or incorrect in 119 cases. Assuming that the 132 cases not examined would have had a similar proportion of problems, an estimated 105 additional cases of correctable data might have been discovered.

For four categories of discrepancies between manual and automated methods, the authors found that the automated method had obtained "correct" information about the patient that was missed during manual CSI case abstraction. These included cases in which (1) the manual coders incorrectly estimated vital signs from HELP system-printed graphic sheets; (2) the chart for some reason did not contain pages normally printed by the HELP system; (3) an item in the electronic database did not appear on the IHELP system paper printout for some reason; and (4) an error occurred in the manual coder. These problems suggested that some of the disagreement between automated and manual methods was due to the manual score being incorrect. Until this point in the study, the manual CSI score had been considered the gold standard for CSI scoring. The authors had planned on comparing the automated score with the manual score and did not expect to find so many cases in which the manual score was in error.

Even when the manual and automated scores agreed, it was not clear that such consensus indicated a "correct" score for either or both methods. Perhaps errors in the manually scored cases were offset by missing data in the automatically scored cases, producing agreement of the two scores but with both scores incorrect. Similarly, one system may have actually determined the "correct" score, but errors in the other system might have caused a false agreement by increasing or decreasing the score that would have been generated by the method without such errors. From random reviews of cases in which the automated and manual scores agreed, the authors estimated that "cor-

![Figure 2](https://academic.oup.com/jamia/article-abstract/3/5/349/689921/2)

Figure 2 Automated score lower than manual score: source of disagreement in 136 discrepancies from 99 cases (Maximum period).

![Figure 3](https://academic.oup.com/jamia/article-abstract/3/5/349/689921/3)

Figure 3 Automated score higher than manual score: source of disagreement in 169 discrepancies from 47 cases (Maximum period).
Coma scores entered into the computer by nurses, di-serum chemistry results, blood gas values, Glasgow Coma Scores entered into the computer by nurses, di-serum chemistry results, blood gas values, Glasgow Sensitive Severity Scoring System (DTS4). Based on electronic data diagnoses on the computerized problem list, and the computerized record of procedures done during the patient's current hospitalization. As in the present study, Fitzpatrick found scoring discrepancies due to patient's current hospitalization. As in the present study, Fitzpatrick found scoring discrepancies due to the updated manual score was "correct."*" Fitzpatrick et al.* reported the use of automatic severity scoring of ICU patients with the Duke Time Sensitive Severity Scoring System (DTS4). Based on the APACHE II scoring system, DTS4 uses vital signs, serum chemistry results, blood gas values, Glasgow Coma Scores entered into the computer by nurses, diagnoses on the computerized problem list, and the computerized record of procedures done during the patient's current hospitalization. As in the present study, Fitzpatrick found scoring discrepancies due to electronic data not appearing in the paper chart, as well as data in the paper chart not found in an electronic form. In addition, the Duke group found, as did the authors of the current study, that the chart abstracts made errors in entering and interpreting data found in the paper chart.

In the authors' environment, automated CSI missed some indicators that led to a higher severity index level because HELP PTXT codes for those concepts have not yet been defined (two out of three missed indicators) and because defined PTXT codes were not used by the nurses (one out of three missed indicators). The authors were not surprised by the number of instances in which existing PTXT codes were not used. In mapping the abstract CSI concepts to the relatively atomic findings in the HELP system, the authors had to be conservative and not interpret too much from the PTXT codes used by the nurses. Most of these instances concerned highly specific parts of the physical examination that may not be noted by all nurses on all shifts. Description of the physical examination is subjective and varies greatly from physician to physician, nurse to nurse, and physician to nurse. Imprecision in clinical description makes mapping the two different vocabularies inexact at best.

This study evaluated an automated interface between a proprietary hospital information system and a proprietary severity index. The results obtained reflect the specific attributes of the two systems. It would be difficult to reproduce this study at many other hospitals because few other hospitals have as many electronic data for patients on multiple nursing floors. The lesson for LDS Hospital and other institutions is that the major limitation of automated clinical severity indexes is the hospital's information system; the depth of concepts represented by the underlying medical vocabulary and the breadth of distribution of the information system throughout the floors of the hospital determine the success of automated severity scoring.

In the authors' environment, automated CSI missed some indicators that led to a higher severity index level because HELP PTXT codes for those concepts have not yet been defined (two out of three missed indicators) and because defined PTXT codes were not used by the nurses (one out of three missed indicators). The authors were not surprised by the number of instances in which existing PTXT codes were not used. In mapping the abstract CSI concepts to the relatively atomic findings in the HELP system, the authors had to be conservative and not interpret too much from the PTXT codes used by the nurses. Most of these instances concerned highly specific parts of the physical examination that may not be noted by all nurses on all shifts. Description of the physical examination is subjective and varies greatly from physician to physician, nurse to nurse, and physician to nurse. Imprecision in clinical description makes mapping the two different vocabularies inexact at best.

This study evaluated an automated interface between a proprietary hospital information system and a proprietary severity index. The results obtained reflect the specific attributes of the two systems. It would be difficult to reproduce this study at many other hospitals because few other hospitals have as many electronic data for patients on multiple nursing floors. The lesson for LDS Hospital and other institutions is that the major limitation of automated clinical severity indexes is the hospital's information system; the depth of concepts represented by the underlying medical vocabulary and the breadth of distribution of the information system throughout the floors of the hospital determine the success of automated severity scoring.

In the authors' environment, automated CSI missed some indicators that led to a higher severity index level because HELP PTXT codes for those concepts have not yet been defined (two out of three missed indicators) and because defined PTXT codes were not used by the nurses (one out of three missed indicators). The authors were not surprised by the number of instances in which existing PTXT codes were not used. In mapping the abstract CSI concepts to the relatively atomic findings in the HELP system, the authors had to be conservative and not interpret too much from the PTXT codes used by the nurses. Most of these instances concerned highly specific parts of the physical examination that may not be noted by all nurses on all shifts. Description of the physical examination is subjective and varies greatly from physician to physician, nurse to nurse, and physician to nurse. Imprecision in clinical description makes mapping the two different vocabularies inexact at best.

This study evaluated an automated interface between a proprietary hospital information system and a proprietary severity index. The results obtained reflect the specific attributes of the two systems. It would be difficult to reproduce this study at many other hospitals because few other hospitals have as many electronic data for patients on multiple nursing floors. The lesson for LDS Hospital and other institutions is that the major limitation of automated clinical severity indexes is the hospital's information system; the depth of concepts represented by the underlying medical vocabulary and the breadth of distribution of the information system throughout the floors of the hospital determine the success of automated severity scoring.

In the authors' environment, automated CSI missed some indicators that led to a higher severity index level because HELP PTXT codes for those concepts have not yet been defined (two out of three missed indicators) and because defined PTXT codes were not used by the nurses (one out of three missed indicators). The authors were not surprised by the number of instances in which existing PTXT codes were not used. In mapping the abstract CSI concepts to the relatively atomic findings in the HELP system, the authors had to be conservative and not interpret too much from the PTXT codes used by the nurses. Most of these instances concerned highly specific parts of the physical examination that may not be noted by all nurses on all shifts. Description of the physical examination is subjective and varies greatly from physician to physician, nurse to nurse, and physician to nurse. Imprecision in clinical description makes mapping the two different vocabularies inexact at best.

This study evaluated an automated interface between a proprietary hospital information system and a proprietary severity index. The results obtained reflect the specific attributes of the two systems. It would be difficult to reproduce this study at many other hospitals because few other hospitals have as many electronic data for patients on multiple nursing floors. The lesson for LDS Hospital and other institutions is that the major limitation of automated clinical severity indexes is the hospital's information system; the depth of concepts represented by the underlying medical vocabulary and the breadth of distribution of the information system throughout the floors of the hospital determine the success of automated severity scoring.

In the authors' environment, automated CSI missed some indicators that led to a higher severity index level because HELP PTXT codes for those concepts have not yet been defined (two out of three missed indicators) and because defined PTXT codes were not used by the nurses (one out of three missed indicators). The authors were not surprised by the number of instances in which existing PTXT codes were not used. In mapping the abstract CSI concepts to the relatively atomic findings in the HELP system, the authors had to be conservative and not interpret too much from the PTXT codes used by the nurses. Most of these instances concerned highly specific parts of the physical examination that may not be noted by all nurses on all shifts. Description of the physical examination is subjective and varies greatly from physician to physician, nurse to nurse, and physician to nurse. Imprecision in clinical description makes mapping the two different vocabularies inexact at best.

This study evaluated an automated interface between a proprietary hospital information system and a proprietary severity index. The results obtained reflect the specific attributes of the two systems. It would be difficult to reproduce this study at many other hospitals because few other hospitals have as many electronic data for patients on multiple nursing floors. The lesson for LDS Hospital and other institutions is that the major limitation of automated clinical severity indexes is the hospital's information system; the depth of concepts represented by the underlying medical vocabulary and the breadth of distribution of the information system throughout the floors of the hospital determine the success of automated severity scoring.

In the authors' environment, automated CSI missed some indicators that led to a higher severity index level because HELP PTXT codes for those concepts have not yet been defined (two out of three missed indicators) and because defined PTXT codes were not used by the nurses (one out of three missed indicators). The authors were not surprised by the number of instances in which existing PTXT codes were not used. In mapping the abstract CSI concepts to the relatively atomic findings in the HELP system, the authors had to be conservative and not interpret too much from the PTXT codes used by the nurses. Most of these instances concerned highly specific parts of the physical examination that may not be noted by all nurses on all shifts. Description of the physical examination is subjective and varies greatly from physician to physician, nurse to nurse, and physician to nurse. Imprecision in clinical description makes mapping the two different vocabularies inexact at best.

The authors thank Blair C. James, software author of the Computerized Severity Index, for his conceptual leadership and programming assistance in this project. The authors are also grateful to Lane Stevens for programming help and to Claudia Madrid for CSI case scoring and advice. The authors thank Dr. Randolph Miller and three anonymous reviewers for substantial editorial efforts. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the National Library of Medicine.

Susan D. Horn, PhD, is founder and owner of International Severity Information Systems (ISIS), Inc., which licenses the Computerized Severity Index (CSI). CSI is a trademark of ISIS.

References


