Research Paper ■

Assessing the Quality of Clinical Data in a Computer-based Record for Calculating the Pneumonia Severity Index

DOMINIK ARONSKY, MD, PETER J. HAUG, MD

Abstract **Objective:** This study examined whether clinical data routinely available in a computerized patient record (CPR) can be used to drive a complex guideline that supports physicians in real time and at the point of care in assessing the risk of mortality for patients with community-acquired pneumonia.

Setting: Emergency department of a tertiary-care hospital.

Design: Retrospective analysis with medical chart review.

Patients: All 241 inpatients during a 17-month period (Jun 1995 to Nov 1996) who presented to the emergency department and had a primary discharge diagnosis of community-acquired pneumonia.

Methods/Main Outcome Measures: The 20 guideline variables were extracted from the CPR (HELP System) and the paper chart. The risk score and the risk class of the Pneumonia Severity Index were computed using data from the CPR alone and from a reference standard of all data available in the paper chart and the CPR at the time of the emergency department encounters. Availability and concordance were quantified to determine data quality. The type and cause of errors were analyzed depending on the source and format of the clinical variables.

Results: Of the 20 guideline variables, 12 variables were required to be present for every computer-charted emergency department patient, seven variables were required for selected patients only, and one variable was not typically available in the HELP System during a patient's encounter. The risk class was identical for 86.7 percent of the patients. The majority of patients with different risk classes were assigned too low a risk class. The risk scores were identical for 72.1 percent of the patients. The average availability was 0.99 for the data elements that were required to be present and 0.79 for the data elements that were not required to be present. The average concordance was 0.98 when all a patient's variables were taken into account. The cause of error was attributed to the nurse charting in 77 percent of the cases and to the computerized evaluation in 23 percent. The type of error originated from the free-text fields in 64 percent, from coded fields in 21 percent, from vital signs in 14 percent, and from laboratory results in 1 percent.

Conclusion: From a clinical perspective, the current level of data quality in the HELP System supports the automation and the prospective evaluation of the Pneumonia Severity Index as a computerized decision support tool.

■ **JAMIA.** 2000;7:55–65.

Affiliation of the authors: LDS Hospital, University of Utah, Salt Lake City, Utah.

The first author was supported in part by grant 31-44406.95 from the Swiss National Science Foundation.

Correspondence and reprints: Dominik Aronsky, MD, Department of Medical Informatics, LDS Hospital, 8th Avenue and C Street, Salt Lake City, UT 84143; e-mail: ⟨lddarons@ihc.com⟩.

Received for publication: 5/19/99; accepted for publication: 9/15/99.

Disseminating and implementing paper-based guidelines in everyday practice remains a major challenge.[1–4] A common reason for the reluctance to use the guidelines is the time required to complete them. Especially in hospitals with a computerized patient record (CPR), paper-based guidelines represent a duplication of data collection that should be avoided.[5] Computerizing a guideline is an attractive and effective means of avoiding the duplication of data collection and the time-consuming manual completion of guidelines.[6] Clinical information systems have the potential to drive guidelines[7,8] and minimize or eventually eliminate additional data collection from the health care providers. Few guideline-based decision support systems are integrated into existing CPRs, support clinicians in real time, and do not require additional data collection.[8–11] One reason for the sparsity of real-time computerized guidelines that do not require additional data entry is that clinically relevant data are not sufficient[12] or are not represented in an easily retrievable and computable format.[5]

A successful approach to computerizing guidelines is the capturing of essential guideline data on a structured encounter screen.[13,14] Guidelines with specifically designed encounter screens capture data in a computable format and have demonstrated both an improvement in documentation[14] and a positive effect on patient outcomes.[15] However, as the guideline increases in complexity, the time required to enter data grows. In contrast, the complexity of a guideline is often necessary to deliver recommendations that are patient-specific. An alternative and desirable approach to delivering patient-specific advice is to integrate a guideline into an existing CPR, taking advantage of the available data.[16,17] This approach may sacrifice data quality because the variables are not specifically collected for driving a guideline but rather are captured for documenting routine patient care.

Even if all the required data elements of a guideline exist in a CPR, the representation format affects the accuracy of guideline recommendations. Data quality needs to be assessed prior to implementation, because erroneous recommendations based on inaccurate data influence guideline acceptance and may raise liability issues.[18] Data quality is a fundamental issue when guidelines are integrated into a CPR. Only a few studies, however, focus on assessing the data quality or accuracy in a CPR.[19,20]

The objective of this study was to determine the data quality of variables routinely collected in an emergency department in driving the Pneumonia Severity Index (PSI).[21] The PSI guideline allows emergency department physicians to assess the risk of mortality in patients who have community-acquired pneumonia. Although the present form of the guideline assures patient-specific recommendations, the complexity remains too high to be easily memorized by clinicians. Computerizing and integrating the guideline into a CPR with verified levels of data quality is a desirable method to deliver real-time support.[22] Our goal was to test whether data routinely available during a patient's encounter in the emergency department can be used to evaluate the PSI and provide physicians with real-time decision support for the management of pneumonia patients at the point of care.

Methods

Following the methodological recommendations of Hogan and Wagner,[20] we describe the setting and charting process in detail.

Setting

LDS Hospital is a 520-bed tertiary-care and university teaching hospital in Salt Lake City, Utah. During the study period, the emergency department staff included 12 full-time, board-certified emergency department physicians. Because of personnel turnover, the nursing staff changed during the study period. There were, generally, 12 full-time and 30 part-time nurses. The emergency department staff cares for more than 25,000 patients per year and uses the HELP (Health Evaluation through Logical Processing) System for data recording and reporting.[23,24] In the emergency department there are 24 HELP terminals, the majority residing in patients' rooms.

The HELP System (version 15) is an inpatient CPR that has a long history and is well known for several integrated decision support systems.[25] It is a commercially available clinical information system (3M HIS, Murray, Utah) that runs on a mirrored Tandem mainframe computer with 12 central processing units. It was developed and is maintained in PAL (PTXT application language), which is a proprietary language of 3M, and in TAL (Tandem application language), which is a proprietary language of Compaq (previously Tandem, located in Houston, Texas). System downtime is 0.15 percent per year.

Defining either the computer-based or the paper-based chart as the official patient record is usually not feasible. Although the HELP System contains most a patient's data, this CPR is complemented with handwritten admission notes, progress notes, and additional forms that reside only in the paper-based record. Similarly, the HELP System contains information that does not enter the paper-based chart unless it is printed and inserted there on request. Thus, the official patient record (the Record) is the combined data from both the CPR and the paper-based chart.

Inclusion and Exclusion Criteria

We included all inpatients who were at least 18 years of age and had a primary discharge ICD-9-CM diagnosis of community-acquired pneumonia (ICD-9 codes 480.0–486.9) during a 17-month period (June 1995 to November 1996).

The chief complaint is a mandatory entry without which the emergency department nurse cannot chart in the CPR. For patients who have cardiac arrest or other life-threatening conditions, nurse charting is performed exclusively on paper and does not enter the CPR. Therefore, we identified patients who had cardiac arrest or another life-threatening condition by the absence of a chief complaint. These patients were excluded from the study. We also excluded patients with pneumonia who should have been admitted directly to the hospital but were admitted first to the emergency department because of a logistical misunderstanding. Because of the misunderstanding, these patients were not seen by an emergency department physician.

The PSI guideline excludes patients who have histories of acquired immunodeficiency syndrome, have a positive titer of HIV antibodies, have been transferred from another acute-care hospital, or have been hospitalized in the seven days prior to the current emergency department encounter. We included patients who met the PSI exclusion criteria, however, because the study did not include the assessment of data quality for variables that evaluate a patient's eligibility criteria for the guideline.

Data Collection in the Emergency Department

The emergency department nurses collect the majority of the data elements required for the PSI. When a patient presents to the emergency department, a registration clerk collects the patient's demographic information. A triage nurse collects the information about chief complaint, current and past histories, and current medication and measures vital signs, including temperature, blood pressure, and heart and respiratory rates. The triage nurse enters the patient's information directly into the HELP System. At the end of the triage, the nurse prints the captured information and attaches the form to the patient's chart. The form is then used by the emergency department physician to record additional findings.

After the patient is transferred to an emergency department room, an assigned nurse charts a more detailed assessment. The assessment is entered in coded form, but the nurse can also chart findings in free text. While the patient is being evaluated, the nurse may measure the vital signs again. Patients whose conditions are considered urgent do not present to the triage nurse but are admitted directly to an emergency department room, where the assigned nurse collects the triage information in addition to the assessment. After evaluating the patient, the emergency department physicians add information to the form printed by the triage nurse.

A clerk orders laboratory tests or radiology examinations through the HELP System. The laboratory results enter the HELP System through an interface with the laboratory computer. The radiologic images are reviewed by the physicians on a radiology workstation in the emergency department. The emergency department physicians review patient information on the HELP System. However, they are not involved in capturing data or entering orders. At the end of the patient's encounter, the nurse discharges the patient from the emergency department and finishes the charting process by recording the discharge time. The physician's dictation and the radiologist's x-ray interpretation enter the HELP System as free-text reports.

The Pneumonia Severity Index Guideline for Community-acquired Pneumonia

The PSI guideline is a severity scoring system that accesses the risk of mortality for pneumonia patients. The PSI was originally developed as a logistic regression[26] but was later converted to a scoring algorithm to ease clinical use and promote the dissemination of the guideline in different settings.[21] The PSI guideline is a two-step algorithm (Figure 1). The first step evaluates 11 variables from the physical examination and the patient's current and past histories. Patients who are 50 years old or younger and have no abnormal findings on the physical examination and the current and past histories are assigned to risk class 1. Otherwise, the patients are assigned to a higher risk class, which is determined in the second step.

The second step evaluates nine additional variables from laboratory tests and the chest x-ray (Table 1). To determine the appropriate risk class in the second step, the physician must first calculate a risk score. The risk score is the sum of points assigned to each of the 20 PSI variables. Finally, the risk class is derived from the risk score and corresponds to a patient's probability of dying (Table 2). The developers of the PSI have suggested that the risk class can be applied as an admission criteria.[21] Patients at low risk of dying might be managed as outpatients, whereas patients at high risk should be admitted to the hospital.

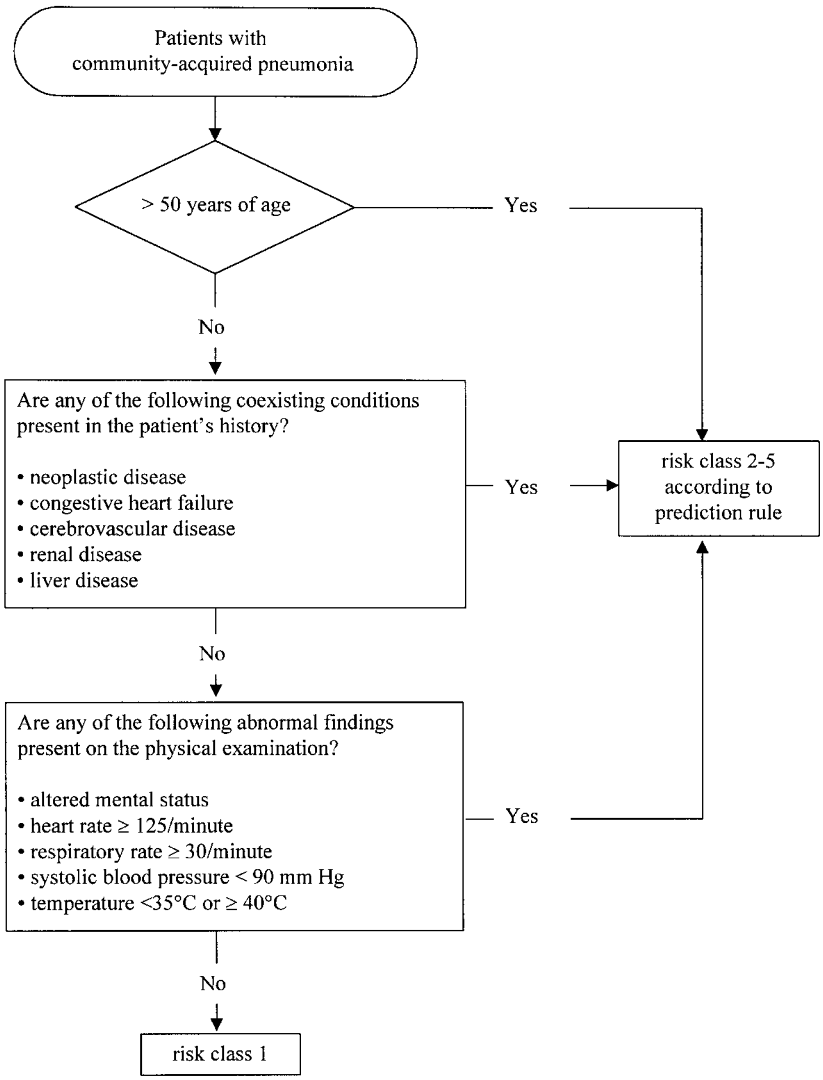

Figure 1 First step of the Pneumonia Severity Index (PSI) scoring algorithm. On the basis of the patient's historical data and vital signs, the clinician assigns the patient to either risk class 1 or a higher risk class. For patients assigned to a higher risk class, a risk score is calculated and used by the clinician to determine the appropriate risk class. (Reprinted with permission from Fine et al.[21] © 1997, New England Journal of Medicine.)

Evaluating the Pneumonia Severity Index

For each patient we calculated the PSI risk score and risk class using data that were available in the CPR during the patient's encounter in the emergency department. The PSI risk score and risk class computed from the CPR data were then compared with the risk measures that were computed using data that originated in the patient's Record (the combined CPR and paper-based record). All information that was actually available in any format during the patient's encounter in the emergency department and that originated in the patient's Record represented our reference standard. The reference standard corresponds to the best information available while the patient was in the emergency department.

For all patients in the study we retrieved the PSI parameters from the CPR through database queries. To assess the five variables of the PSI that involve coexisting conditions (neoplastic disease, congestive heart failure, cerebrovascular disease, renal disease, and liver disease), we constructed a list of terms representing each coexisting condition. The list of terms was compiled by review of the free-text fields of the nurse-charting entries in the CPR for patients who were seen in the six months following the study period (Dec 1996 to May 1997). For patients in the study we inferred the presence of disease if one of the terms was present in the free-text field of the current history, the past history, or the current medication. We considered only patient information that was recorded while the patient was in the emergency department. The emergency department encounter started at the time a patient was registered by either the registration clerk or the emergency department nurse and ended at the time the patient was admitted to the hospital.

Table 1 ■

Variables Evaluated in the Second Step of the Pneumonia Severity Index (PSI) Scoring Algorithm

Patient Characteristic	Source of Data in the HELP System	Type of Data in the HELP System	Points Assigned for Abnormality
Demographics:			
Age:			
Men	Registration	Coded	Age (year)
Women	Registration	Coded	Age (year) −10
Nursing home	Triage	Free text	+10
Coexisting illnesses:			
Neoplastic disease	Triage	Free text	+30
Liver disease	Triage	Free text	+20
Congestive heart failure	Triage	Free text	+10
Cerebrovascular disease	Triage	Free text	+10
Renal disease	Triage	Free text	+10
Physical examination findings:			
Altered mental status	Nurse assessment	Coded	+20
Respiratory rate	Triage	Numeric	+20
Systolic blood pressure	Triage	Numeric	+20
Temperature	Triage	Numeric	+15
Heart rate	Triage	Numeric	+10
Laboratory findings:			
Arterial pH	Laboratory	Numeric	+30
Blood urea nitrogen	Laboratory	Numeric	+20
Sodium	Laboratory	Numeric	+20
Glucose	Laboratory	Numeric	+10
Hematocrit	Laboratory	Numeric	+10
p_aO_2 or SpO_2	Laboratory or triage	Numeric	+10
Radiographic finding:			
Pleural effusion	ED physician's report	Free text	+10

NOTE: The second step of the PSI scoring algorithm evaluates 20 variables to establish a risk score for the patient. In the HELP system the variables differ in format and have different sources. The score for the PSI is calculated by adding the patient's age and the points assigned for each abnormal finding. ED indicates emergency department.

Table 2 ■

Association of Risk Score with Risk Class and Mortality in the Pneumonia Severity Index (PSI) Scoring Algorithm

Risk Score	Risk Class	Mortality (%)
Based on step one	1	<0.5
≤70	2	0.5–0.9
71–90	3	1–3.9
91–130	4	4–10
>130	5	>10

NOTE: The second step of the PSI scoring algorithm assigns the patient to a risk class based on the calculated risk score. The assigned risk class corresponds to a probability of death.

For every patient in the study we obtained the paper-based charts from the medical records department. To abstract data from the patient's record, a self-coding data sheet was completed. The self-coding data sheet has been previously used in our institution for collecting data from the charts of pneumonia patients. As a safeguard against introducing bias into the unblinded review process, the records of 24 patients (10 percent) were randomly selected and were re-evaluated after a two-month interval.

Evidence of abnormal findings in the CPR or the Record that became available after the patient had left the emergency department was not considered for the evaluation of the PSI. For example, if a blood gas analysis was in process but the results were not available while the patient was present in the emergency department, the results were not included in the computation of the PSI. If a series of data elements from the same category was present, only the first measurement, and not the most abnormal measurement, was considered. For example, if the patient's initial systolic blood pressure was 100 mmHg, but subsequent measurements fell below the critical value of the PSI because the patient's condition worsened, only the initial value of 100 mmHg was included in calculating the risk score.

During the study period, the emergency department physician's interpretation of chest x-rays was not available, because information from the radiology database could not be accessed in real time. However, the emergency department physician's interpretation became available for real-time evaluation after the study period. A current implementation of the PSI would include the emergency department physician's interpretation of chest x-rays; consequently, we included the variable "pleural effusion" in the evaluation of the PSI. The principal reason for the emergency department physicians' real-time documentation of their chest x-ray interpretations is to facilitate fast and successful communication with the radiologists. To determine whether a pleural effusion was present, we manually retrieved the emergency department physician's chest x-ray findings from the emergency department physician's free-text report. Discrepancies between the emergency department physician's interpretation of the chest x-ray and that of the radiologist were not considered, because the radiologist's interpretation is not available during the patient's encounter.

Outcome Measures

The risk class is the clinically relevant measure that provides the emergency department physician with objective information about the patient's risk of mortality. However, for evaluating the impact of data quality of the CPR on the PSI, we determined the differences for both the risk scores and the risk classes, because an aberrant value might influence the patient's risk score without changing the risk class.

We assessed the data quality of the 20 PSI guideline variables along the two dimensions "availability" and "concordance." *Availability* is the proportion of observations from the reference standard that were actually recorded in the CPR. *Concordance* is the proportion of observations that are identically recorded in the CPR and the reference standard.

Because the emergency department created different standard sets of clinical variables that are required to be collected during a patient's encounter, we distinguished between the availabilities of required and optional PSI variables. The standard sets depend on the patient's chief complaint and triage category. The triage category of patients with suspected pneumonia is commonly "nonurgent." For nonurgent patients, the standard set of variables required for computing the PSI includes age, gender, the current and past history, the current medication, the systolic blood pressure, the heart and respiratory rate, and the temperature. The standard data set covers 12 PSI variables that are required to be present. The remaining eight PSI variables are not part of the standard data set and are, therefore, not required to be collected for every pneumonia patient. For example, a blood gas analysis does not represent an indispensable test for the diagnosis or management of pneumonia. Accordingly, "arterial pH" represents an optional PSI variable that is not required to be collected.

We quantified concordance for all variables of the PSI except "pleural effusion," because the data source for pleural effusion was identical for both the input variable and the reference standard. Concordance was assessed both when an individual patient was the unit of analysis and when all the variables were considered as one complete set. To examine the discordant data elements in more detail, we determined the types of error and causes of error. To explore the types of error, we stratified the errors into four categories, which depend on the data format and the source of the variables—free text, coded data, laboratory data, and vital signs. To examine the causes of error, we stratified the errors into two categories. One category included errors that were attributable to the emergency department charting process, and one category contained errors that originated from the computerized evaluation of the PSI algorithm. For example, if the emergency department nurse did not record that the patient had congestive heart failure or renal disease, the omission was categorized as an emergency department charting error. If the emergency department nurse misspelled a term, the error was categorized as a system error due to computerization of the PSI, because a parsing algorithm should be able to detect common misspellings. Distinguishing between different sources of errors means determining whether the errors are related to the actual computerization of the guideline or to the emergency department charting process in general. For example, if emergency department physicians do not document the presence of a pleural effusion, the resulting error is unrelated to the computerization of the guideline. Determining different sources of errors supports future efforts to improve the data quality for driving the PSI guideline.

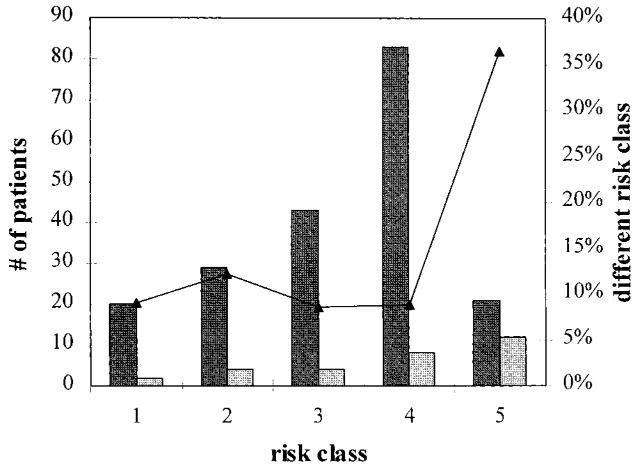

Figure 2 Analysis of risk classes. The chart shows patients' true risk classes as derived from the reference standard. For each risk class, the number of patients with identical risk classes (dark shading) and the number with different risk classes (light shading) are graphed. Solid triangles indicate the percentage of patients assigned to different classes. A different risk class was derived from the CPR data for 20 patients (13.3 percent). The proportion of misclassification was greatest in risk class 5.

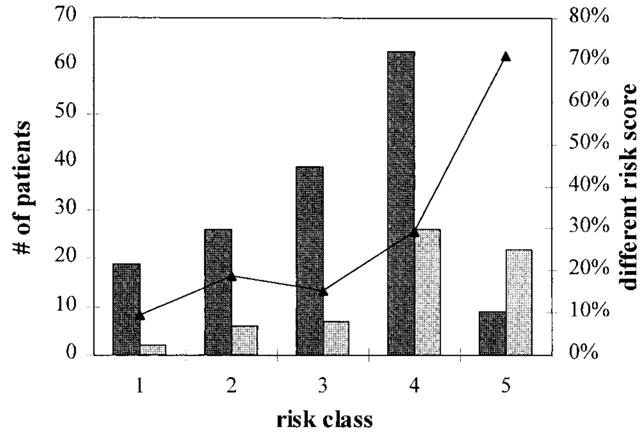

Figure 3 Analysis of risk scores by risk class. The chart shows patients' true risk classes as derived from the reference standard. For each risk class, the number of patients with identical risk scores (dark shading) and the number with different risk scores (light shading) are graphed. Solid triangles indicate the percentage of patients assigned to different classes. An identical risk score was derived from the CPR data for 163 patients (72.1 percent). In risk classes 4 and 5, the proportion of incorrect scores was greatest.

Results

In the 17-month study period, 226 of 241 inpatients met the inclusion criteria. Of the 15 excluded patients, 13 patients had a missing chief complaint, 1 patient was erroneously admitted to the emergency department without being seen by an emergency department physician, and 1 patient had both a missing chief complaint and was admitted erroneously to the emergency department instead of directly to the ward. The second chart audit revealed two discrepancies, when a coexisting disease was missed. The intrarater agreement for assessment of a patient's risk score was 92 percent.

The risk class computed for an individual patient from the CPR and from the Record was identical for 196 patients (86.7 percent). Among the remaining 30 patients (13.3 percent), the CPR-derived risk class was two risk classes lower for 1 patient (3 percent), one risk class lower for 22 patients (73 percent), and one risk class higher for 7 patients (20 percent). For one patient (3 percent), the CPR-derived risk class was the same because simultaneous errors (committed and omitted) equalized the two risk scores and the resulting risk classes. Twenty (87 percent) of the 23 patients with a lower CPR-derived risk class were assigned risk class 3, 4, or 5. Figure 2 summarizes the overall misclassification of the CPR-derived risk class stratification.

The risk score obtained from the CPR was identical for 163 patients (72.1 percent) and different for 63 patients (27.9 percent). Of the 63 patients with a different risk score, the score was lower for 50 patients (79 percent), equal for 1 patient (3 percent), and higher for 12 patients (19 percent). For the patient with an equal risk score, the points attributed to various errors resulted in an equivalent risk score. For patients who had a false risk score computed, a total of 78 errors occurred. A single error occurred for 52 patients, two errors for 8 patients, three errors for 2 patients, and four errors for 1 patient. Among 63 patients assigned risk class 3, 4, or 5, errors affected the risk score of 55 (87 percent) and accounted for 70 (90 percent) of all 78 errors. Figure 3 summarizes the overall scoring errors of the CPR stratified by risk class.

Among the 63 patients with a different risk score, a change of the risk class occurred for 29 (46 percent), of whom 6 patients had a higher risk class and 23 had a lower risk class derived from the CPR. For the remaining 34 patients, the deviant risk score did not influence the risk class. Of the 29 patients with a different CPR-derived risk class, 21 patients were in risk class 4 or 5.

The average availability was 0.991 for the 12 PSI data elements that were part of the standard data set and were required to be present. The number of missing values and the availability for each required variable

Table 3 ■

Data Availability and Number of Missing Entries for Pneumonia Severity Index (PSI) Variables of the Standard Data Set for 226 Patients

Characteristic	Missing Values	Data Availability
Age	0	1.000
Gender	0	1.000
Current history	1	0.996
Past history	0	1.000
Medication	6	0.973
Heart rate	3	0.987
Respiratory rate	1	0.996
Systolic blood pressure	4	0.982
Temperature	3	0.987
Average		0.991

NOTE: The nine required variables of the standard data set cover 12 data elements of the PSI.

Table 4 ■

Data Availability and Number of Missing Entries for Pneumonia Severity Index (PSI) Variables That Are Not Part of the Standard Data Set for 226 Patients

Characteristic	Missing Values	Data Availability
Oxygen saturation	23	0.898
Mental status	37	0.836
Blood urea nitrogen	14	0.938
Sodium	8	0.965
Glucose	14	0.938
Hematocrit	9	0.960
Arterial pH	84	0.628
Pleural effusion	195	0.137
Average		0.788

NOTE: Although the clinical parameters are needed for the evaluation of the PSI, the variables are not obtained for every pneumonia patient as part of clinical care in the emergency department.

are shown in Table 3. For the eight variables that were not part of the standard data set, the average availability was 0.788 for the 226 patients. The number of missing values and the data availability are shown in Table 4. Arterial pH and pleural effusion were the variables with the lowest data availability (0.63 and 0.14, respectively). Blood gas analysis was not performed for 84 patients, and the presence of a pleural effusion was noted for 31 patients. The data availability for the four optional PSI variables that were results of laboratory tests was 0.94 or higher. Missing data variables yielded an error in the risk scores of 17 patients (7.5 percent).

The concordance of data variables was assessed without the finding "pleural effusion." The average concordance for the remaining 19 PSI variables in the 226 patients was 0.982 (4,216 concordant characteristics divided by the total number of 4,294 characteristics). For the 19 considered variables, the average number of concordant variables per patient was 18.65.

Different risk scores originating from the free-text fields accounted for 50 errors (64 percent), of which 34 were attributed to the emergency department nurse charting and 16 to the parsing algorithm. The 34 free-text errors occurred because a PSI-relevant coexisting disease was not charted for 22 patients, and 12 patients were not identified as coming from a nursing home. The 16 errors caused by an imprecise parsing algorithm were due to misspellings for 3 patients, an incomplete and imperfect list of query terms for 12 patients, and the misspelling of a phrase that simultaneously represented a missing query term for one patient. Different risk scores caused by errors in the coded data element ("mental status," "gender") accounted for 16 errors (21 percent), all of which were ascribed to "mental status" and none to "gender." All the errors in coded data were attributed to the emergency department nurse. One pathologic laboratory finding (1 percent) was missed because the patient failed outpatient treatment for pneumonia and was admitted to the hospital the following day with an abnormal blood urea nitrogen value. Because the patient's type of encounter was converted from outpatient to inpatient registration, the abnormal blood urea nitrogen value was the second measurement for the patient's encounter. Among nine patients, there were 11 errors (14 percent) in the vital signs, of which

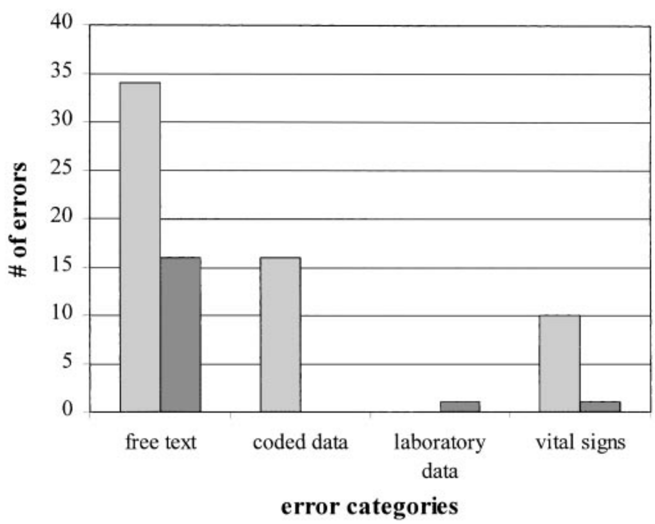

Figure 4 Analysis of types of errors and their causes. The most prevalent type of error occurred in the free-text category and accounted for 64 percent of all errors. The laboratory data had the most correct data. The cause of error was attributed to emergency department nurse charting (light shading) for 77 percent of errors and to system errors (dark shading) for the rest.

10 were attributable to the emergency department nurse and one was caused by a rounding error. For six patients the emergency department nurse did not chart the oxygen saturation under room air condition, and for two patients no vital signs were charted, which accounted for four missing values.

Errors attributable to emergency department nurse charting accounted for 77 percent of the total, and the remaining 23 percent were categorized as system errors. Figure 4 summarizes the type and cause of errors that resulted in a different PSI risk stratification.

Discussion

This study investigated the quality of data routinely available in the HELP System to drive the PSI guideline in real time and to deliver decision support at the point of care. Although the PSI was developed following rigorous methodologic requirements,[22,27] the complexity remains high, jeopardizing its dissemination and implementation. Despite its complexity, the PSI algorithm fulfills the criteria for a successful computerization, including the presence of clear definitions of variables and a decidable algorithm.[28] However, a CPR that accommodates the PSI guideline should meet additional criteria, such as the presence of sufficient and routinely available data and a high level of data quality.

We analyzed data quality from the clinical perspective on computerizing the severity index. The clinically relevant piece of information that is finally presented to the clinician is the patient's risk class. An identical risk class was obtained from the CPR for 86.7 percent of patients. This result is encouraging, considering that the paper-based version of the guideline is too complex to be easily memorized, many data variables originate from nurse charting, no additional data entry was required to achieve this level of data quality, and information from free-text fields was included. Presenting the CPR-available PSI information at the time of decision making enables clinicians to improve the accuracy of the risk class by correcting errors or adding missing information.

The availability of the PSI variables included in the standard data set is sufficiently high to run the respective part of the PSI from the CPR. In one third of the 18 incomplete records, the emergency department nurse left the entry for "current medication" empty. The emergency department nurse is supposed to enter "none" in a free-text field if the patient does not report any pertinent information. A charting practice that leaves free-text fields blank introduces ambiguity in the interpretation, because it is not known whether the patient is actually not taking any medication or whether the emergency department nurse forgot to chart the information or even ask the patient for it.

The interpretation of the data availability for the optional PSI variables is difficult because the data were evaluated retrospectively and were compared with information in the patient's record. If the emergency department physicians choose to apply the PSI guideline routinely, they would be required to obtain an oxygen saturation, laboratory data, and a blood gas analysis for every pneumonia patient who did not fall into the lowest risk class. In a prospective analysis of the PSI, we would predict that the data availability of the optional PSI variables would be even higher.

The analysis of the risk score presents a different perspective and is a direct result of the data quality. Although the average concordance of 0.98 for variables that are part of the standard data set appears to be high, the 78 errors considerably affected the risk scores. More than one quarter (27.9 percent) of the CPR-derived risk scores were different and underestimated the real risk for most patients. The underestimation of the CPR-derived risk score raises a problem for a computerized implementation. The calculated risk score might translate into a lower risk class, underestimating the patient's true risk of mortality. If the risk class is used as an admission criterion,[21] risk classes that are too low mean that more patients are treated as outpatients when they actually should be admitted to the hospital.

The most frequent errors occurred in the free-text category, and the most common cause of these errors was inaccurate nurse charting. Errors originating from the free-text fields might be reduced by applying a more sophisticated parsing algorithm, especially as natural language processing methods become available and are incorporated into CPRs. Free-text fields remain difficult to use in decision support systems. An alternative approach to increasing data quality is to try to encode the terms that appear most frequently in the free-text fields.

The interpretation of the risk scores, however, must be viewed in terms of the conservative approach of the study design. Of the 372 inpatients and outpatients who were diagnosed with pneumonia in the emergency department during the study period, we included only the 241 inpatients. In general, patients admitted to the hospital have more coexisting diseases and more abnormal findings than outpatients. To evaluate the PSI computerization, we focused on inpatients only, because outpatients present with few abnormal findings. The computerized records of inpatients provide more opportunities for the commission of errors in the computerization process than do the records of outpatients.

It is surprising that all 20 variables needed to compute the PSI during a patient's encounter in the emergency department are available in electronic form which promotes an automation of the PSI. However, it is important that all the variables of the PSI or of any other guideline are present in computable format and do not require additional data input. The emergency department is a busy clinical setting for the implementation of guidelines,[4] emphasizing the importance of complete automation of the PSI.

A limitation of this study is that one author performed both the CPR evaluation and the review of the patient's Record. Ideally, persons blinded to the purpose of the study would carry out both the CPR evaluation and the review of Records. We attempted to minimize observer bias by standardizing the data collection with a self-coding data sheet and by performing a second chart audit of 10 percent of randomly selected patient Records.

Another limitation concerns the vital signs. For clinical ease, the PSI dichotomizes the continuous values such as heart rate or oxygen saturation into normal and abnormal values. Therefore, we determined whether the actual value was identical not on a continuous scale but on a dichotomous scale. For example, a heart rate value was considered normal even if a data entry error occurred when the nurse incorrectly charted the rate at 60 beats/min instead of 80 beats/min. The majority of data entry errors are impossible to detect because, in contrast to the nurse charting notes, the dictated reports do not contain a time stamp indicating when values were recorded. To resolve discrepancies between a normal value obtained from the CPR and an abnormal value obtained from the Record, the abnormal value was considered the correct one. We did not quantify data entry errors that may have occurred when values from both the CPR and the Record were normal. However, data entry errors seem not to be an important cause of incorrect data.[19]

To ease and promote its clinical application, the PSI was simplified from a logistic regression to a scoring algorithm. Computerizing the PSI on the basis of the original logistic regression takes advantage of the computational power of a clinical information system and provides a probability and a 95 percent confidence interval. For clinicians, the probability may represent a more precise and intuitive mortality measure than the less meaningful PSI risk class.

The computerization of paper-based guidelines is desirable, because it assists health care providers with easily accessible decision support at the point and the time of care. The computerized representation of the validated and clinically useful PSI guideline supports the implementation and dissemination of the prediction rule. Although the paper-based PSI guideline has advantageous characteristics for an automation, successful clinical implementation depends on the availability of high-quality data. The level of data quality should be assessed prior to implementation, because identification of the sources of errors supports efforts to improve data capture and provide correct and complete data. Enabling clinicians to review and modify the data variables used to generate guideline suggestions represents a possible approach to achieving an accurate risk measure. This approach comes at the cost of additional data entry, however, and it remains uncertain whether clinicians are willing to trade additional data entry for a higher level of data quality in the CPR. Demonstrating a high level of data quality prior to guideline implementation increases the credibility toward computer-generated guideline recommendations and ensures that clinicians can eliminate existing inaccuracies in the risk assessment with few corrections or additions. Only implementation of the PSI as an integrated computerized decision support tool will indicate whether the automated recommendations will influence the clinician's decision making.

The authors thank Wendy Webber Chapman, Nathan C. Dean, MD, and Charles Lagor, MD, for helpful comments in the preparation of the manuscript.

References ■

1. Elson RB, Connelly DP. Computerized patient records in primary care: their role in mediating guideline-driven physician behavior change. Arch Fam Med. 1995;4:698–705.
2. Zielstorff RD. Online practice guidelines: issues, obstacles, and future prospects. J Am Med Inform Assoc. 1998;5:227–36.
3. Schriger DL. Emergency medicine clinical guidelines: we can make them, but will we use them? Ann Emerg Med. 1996;27:655–7.
4. Berger JT, Rosner F. The ethics of practice guidelines. Arch Intern Med. 1996;156:2051–6.
5. Barnett GO, Winickoff RN. Quality assurance and computer-based patient records. Am J Public Health. 1990;80:527–8.
6. Nilasena DS, Lincoln MJ, Turner CW, et al. Development and implementation of a computer-generated reminder system for diabetes preventive care. Proc Annu Symp Comput Appl Med Care. 1994:831–5.
7. Elson RB, Connelly DT. Applications of computer-based clinical guidelines. JAMA. 1998;279:989–90.
8. Lobach DF, Hammond WE. Computerized decision support based on a clinical practice guideline improves compliance with care standards. Am J Med. 1997;102:89–98.
9. Tierney WM, Overhage JM, Takesue BY, et al. Computerizing guidelines to improve care and patient outcomes: the

example of heart failure. J Am Med Inform Assoc. 1995;2: 316–22.
10. Evans RS, Pestotnik SL, Classen DC, et al. A computer-assisted management program for antibiotics and other anti-infective agents. N Engl J Med. 1998;338:232–8.
11. Morris AH, Wallace CJ, Menlove RL, et al. Randomized clinical trial of pressure-controlled inverse ratio ventilation and extracorporeal CO_2 removal for adult respiratory distress syndrome. Am J Respir Crit Care Med. 1994;149:295–305.
12. Tierney WM, Overhage JM, McDonald CJ. Toward electronic medical records that improve care. Ann Intern Med. 1995;122:725–6.
13. Zielstorff RD, Barnett GO, Fitzmaurice JB, et al. A decision support system for prevention and treatment of pressure ulcers based on AHCPR guidelines. Proc AMIA Annu Fall Symp. 1996:562–6.
14. Schriger DL, Baraff LJ, Rogers WH, Cretin S. Implementation of clinical guidelines using a computer charting system: effect on the initial care of health care workers exposed to body fluids. JAMA. 1997;278:1585–90.
15. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. JAMA. 1998;280:1339–46.
16. Tierney WM, Overhage JM, McDonald CJ. Computerizing guidelines: factors for success. Proc AMIA Annu Fall Symp. 1996:459–62.
17. Classen DC. Clinical decision support systems to improve clinical practice and quality of care. JAMA. 1998;280:1360–1.
18. Jacobson PD. Legal and policy considerations in using clinical practice guidelines. Am J Cardiol. 1997;80(8B):74H–79H.
19. Wagner MM, Hogan WR. The accuracy of medication data in an outpatient electronic medical record. J Am Med Inform Assoc. 1996;3:234–44.
20. Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc. 1997;4:342–55.
21. Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med. 1997;336:243–50.
22. Wasson JH, Sox HC. Clinical prediction rules: have they come of age? JAMA. 1996;275:641–2.
23. Kuperman GJ, Gardner RM, Pryor TA. HELP: A Dynamic Hospital Information System. New York: Springer-Verlag, 1991.
24. Gardner RM, Pryor TA, Warner HR. The HELP hospital information system: update 1998. Int J Med Inf. 1999;54:169–82.
25. Haug PJ, Gardner RM, Evans RS. Hospital-based decision support. In: Berner ES (ed). Clinical Decision Support Systems: Theory and Practice. New York: Springer-Verlag, 1998.
26. Fine MJ, Hanusa BH, Lave JR, et al. Comparison of a disease-specific and a generic severity of illness measure for patients with community-acquired pneumonia. J Gen Intern Med. 1995;10:359–68.
27. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules: applications and methodological standards. N Engl J Med. 1985;313:793–9.
28. McDonald CJ, Overhage JM. Guidelines you can follow and can trust: an ideal and an example. JAMA. 1994;271:872–3.