G. Hripcsak

Center for Medical Informatics,
Columbia-Presbyterian Medical Center,
New York, NY, USA

Synopsis

Health and Clinical Management

As a field matures, research often shifts from topics that are most appealing to investigators (for example, topics of personal interest) to topics that are most important to the community as a whole. Medical informatics is maturing as a field, and the use of medical informatics for health and clinical management demonstrates its emphasis on community concerns. The papers in this section address the needs of the health care provider and the clinical investigator. Any method or system professing to be useful to the community must undergo sufficient evaluation; we see a common theme of evaluation in all the papers.

In the first paper, *Use of MEDLINE by physicians for clinical problem solving* [1], Lindberg and coauthors address the use of an electronic bibliographic database 'MEDLINE' for clinical care. The authors analyzed 1158 reports from 552 health professionals on how their use of MEDLINE affected patient care in terms of positive impact on clinical problem solving and positive impact on patient outcomes. Subjects, who were mostly clinicians and had had varying degrees of experience with MEDLINE, were interviewed via telephone about recent MEDLINE searches. The interview questions were tailored depending on whether the subject felt that a MEDLINE search had been effective. An impressive 86% reported that they considered their searches "effective." Forty-one percent of searches (476) affected patient care. For these latter searches, motivation for the search, effect on clinical problem solving, and effect on outcomes were analyzed. Answers were classified using the Critical Incident Technique. A three-level hierarchy of positive effects on problem solving is shown in the paper's first table. This extensive list may become invaluable for educating health professionals about the many ways that MEDLINE can be exploited for clinical care. Similarly, the paper's second table lists positive impacts on outcomes. Out of 455 reports, 429 showed a positive impact, and none showed a negative impact. While this may have been biased somewhat by the positive phrasing of interview questions, one would expect disgruntled users to speak up. At least one example of a negative impact has been noted at Columbia-Presbyterian Medical Center. A clinician who was treating a diabetic patient with cellulitis was looking for an alternative therapy for the infection. A MEDLINE search produced a study that suggested that a new oral antibiotic was equivalent to a traditional intravenous approach. The patient did not do well on the regimen; and the study's recommendation has not been corroborated by an independent group, nor is it sanctioned by infectious disease experts. These reports are rare, however, and may not be different from reports that could be obtained about the use of textbooks or colleague consultations. Subjects reported that they used MEDLINE because it is up-to-date, because it is readily available, and because of the power of the searches. As search techniques and information presentation improve, we can hope that MEDLINE will rival good textbooks in their two advantages: balanced presentation and synthesized conclusions.

The next three papers take novel approaches to clinical problems. All three are in early development stages, but illustrate uses of medical informatics in clinical care. In *Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery* [2], Tu and Guerriere present a neural network intended to predict the length of stay in an intensive care unit (ICU) for patients who have just undergone cardiac surgery. ICU beds represent a limited resource, and patients who require a prolonged stay can reduce the number of cardiac operations done at a medical center. Being able to predict ICU length of stay in advance would permit more efficient operating room scheduling, better understanding of the patient's prognosis and, potentially, the ability to alter management to reduce the stay (e.g., prophylaxis in a subgroup). The neural network accepts 15 relevant input variables, such as age, type of surgery, and comorbid diseases. These 15 inputs map to 12 intermediate (hidden) units, and these, in turn, map to one output unit, whose value (0 to 1) measures the likelihood that the patient will require more than two days in the ICU. The neural network is a standard back-propagation design with a logistic activation function. The network was trained on a set of 713 cases, and tested on an independent set of 696 cases. Its performance was assessed using the area under its ROC curve, which was 0.6960, but this result was not compared to that of competing methods such as logistic regression or recursive partitioning. Interestingly, the result for the training set was not significantly different than that for the test set, so the network did not overfit the training data. The network output was stratified into three intervals (0-0.25, 0.25-0.5, 0.5-1.0), which corresponded to 16.3%, 35.3%, and 60.8% chance of requiring a long ICU stay. An abstract comparison between neural networks and logistic regression is made in the paper, but a comparison based on real data is needed. Such a comparison would indicate whether the network's ability to handle complex dependencies among the variables is important in this domain.

In the next paper, *A burn patient resuscitation therapy designed by computer simulation (BET). Part 1: simulation studies* [3], Roa and Gomez-Cia demonstrate the use of simulation to design and test therapeutic regimens. They describe a resuscitation therapy for patients with severe burns. The authors have created a control system model that maps three components of fluid therapy (fluid, colloid, and electrolyte) to associated clinical outcomes (plasma protein concentration, hematocrit, arterial pressure, extracellular osmolarity, and diuresis). Based upon this model, the authors have designed a fluid therapy called BET, which minimizes the deviation from the patient's initial values for those parameters. A simulator based upon the model demonstrates that the performance of BET is superior to alternative therapies. As implied by the title ("Part 1"), further evaluation is warranted. That a therapy based upon a model should do well when judged by that model is not surprising, although it is a crucial first step. A similar comparison using an independently developed simulation model would be the next logical step, followed eventually by concrete experiments.

In *Pain assessment with interactive computer animation* [4], Swanston and coauthors describe the use of interactive computer graphics to assess a patient's perception of pain. The standard technique to assess pain in a reproducible fashion is to use established questionnaires augmented with a linear rating scale (printed on paper). The authors created an interactive graphical interface to measure pain, in order to better match the patient's perception of pain and to reduce the reliance on linguistic competence. The system addresses four specific pain categories: pressure pain is represented by a diagram of a vise, burning pain by a red color, throbbing pain by a pounding hammer, and piercing pain by a needle. Overall degree of pain is represented on a simple analog scale, much like the paper counterpart. The patient is asked to pick one or more categories and to indicate the degree that the characteristic is present by manipulating the figure (for example, tightening the vise to represent more pressure pain); the patient then indicates the overall degree of pain on the analog scale. In an evaluation with 54 patients from a pain clinic, the results of the interactive animation were compared to a standard questionnaire. A high correlation was found for overall degree of pain, probably because both methods used a linear scale. The correlation between the computer pain category data and the paper-based measurements was much lower. The authors discovered that patients who chose to represent their pain with more than one category were more likely to have a good correlation between the computer and paper-based approaches. The analysis could be improved in several ways (for example, four patients were dropped from the study because they chose no category, but this is exactly what may happen if the system is really used), and the results do not demonstrate reproducibility, but anecdotes indicate that for at least some patients, the animated method better corresponds to their own perception of pain.

The section concludes with two strong papers about clinical research. In *Discordance of databases designed for claims payment versus clinical informations systems* [5], Jollis and coauthors compare data collected for the purpose of processing insurance claims to clinical data collected during a prospective study of ischemic heart disease. Claims data are represented as clinically modified International Classification of Diseases (ICD-9-CM) codes, and are used to substantiate charges for inpatient health care. The data are abstracted by medical records technicians, who rely on discharge diagnoses, information from progress notes, and test results from hospitalization. The clinical data are collected by cardiology fellows who have available to them history, physical exam, test results, and cardiac catheterization results. Twelve clinical variables were selected for analysis, based upon their importance in coronary artery disease prognosis and ability to be coded in ICD-9-CM. The claims data and clinical records from 12,937 patients were compared. The highest chance-corrected agreement was 0.83 for diabetes mellitus, and the lowest 0.09 for unstable angina. Nine out of twelve had chance-corrected agreement lower than 0.50. The overall agreement for all variables was 0.75. There were 15,678 instances in which the clinical data identified one of four preselected diagnoses missed by the claims data

Sixty such cases were reviewed manually; 54 appeared to be errors in the claims data, and 6 had insufficient information to tell what went wrong. There were 1,276 instances in which the claims data identified a diagnosis that the clinical data missed. Of 60 such cases, only 6 were proven to be errors in the clinical data. Therefore, the clinical data are far more reliable than the claims data; and the claims data are more likely to be missing a diagnosis than to have an incorrect one. Several conclusions can be drawn. Claims data lack the completeness and accuracy necessary to be used reliably in clinical care and research. In many places, only claims data are available; therefore the incompleteness must be accounted for in analyses. Errors during claims coding are unlikely to be random; the resulting bias will be difficult to remove. Furthermore, if claims data are abominable for clinical use, how good are they for claims use (that is, the accurate allocation of resources)? Perhaps we need both better clinical information systems and better claims data.

In the last paper, *A methodology for determining patient's eligibility for clinical trials* [6], Tu and coauthors attempt to facilitate the difficult and time-intensive process of screening patients for clinical trials. Eligibility in a clinical trial is expressed as a set of criteria. For a given patient, a criterion may be true, false, or not known. If criteria were only true or false, then the problem would be straightforward. But criteria that are not known can be inferred, at least probabilistically, from default assumptions (if scleroderma is not mentioned, then it is probably not present), older data (if the platelet count was recently low, then it is probably still low), and related data (if the white count is less than 1000, then the lymphocyte count is less than 1000). The paper describes two approaches to inferring degree of eligibility based on whatever data are available for a patient. The qualitative approach uses heuristic rules and default assumptions to decide the degree to which a patient meets a criterion (meets, probably meets, not known, probably fails, or definitely fails). The criteria are then combined, resulting in a final, qualitative assessment. The second approach uses a Bayesian belief network to represent eligibility criteria. The network contains nodes that represent criteria, propositions (for example, the conjunction of two criteria), and related evidence. Underlying distributions are chosen, and links are modeled. Missing data are then handled automatically through the network. The network produces a probability that the patient qualifies for the clinical trial. This more quantitative result comes at a price: more complex assumptions about distributions, more complex modeling, and potential performance problems. A combined approach avoids some complex modeling of dependencies among criteria, but produces only a qualitative answer. In an evaluation, the data for 60 patients were abstracted manually from medical records and analyzed according to the methodology. Twelve patients who were not actually enrolled in protocols were found to be eligible (or probably eligible) according to the methodology. Unfortunately, the reasons why the patients were not enrolled were not ascertained, and it is unclear whether enrollment was accidentally missed, determined to be inappropriate, or refused by the patient. The authors note that some eligibility criteria are "subjective" in that they involve a physician's judgment (for example, how likely is a patient to comply with a protocol). Subjective criteria are supported by the methodology, although they are not included in the paper's examples. The randomization phase of a clinical trial should eliminate bias due to the use of subjective criteria. But when the results of a trial are published, it is up to the clinician to decide whether her patient is similar enough to the patients in the trial to justify following the trial's recommendations. This decision is only possible if the criteria are clearly stated. If this paper's methodology can codify otherwise vague criteria, then it may perform an even greater service to clinical trials: making it easier to apply results to real patients.

References

1. Lindberg DAB, Siegel ER, Rapp BA, Wallingford KT, and Wilson SR. Use of MEDLINE by physicians for clinical problem solving. JAMA 1993;269:3124-9.
2. Tu JV, and Guerriere RJ. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. Comput Biomed Res 1993;26:220-9.
3. Roa L, and Gomez-Cia T. A burn patient resuscitation therapy designed by computer simulation (BET). Part 1: simulation studies. Burns 1993;19:324-31.
4. Swanston M, Abraham C, Macrae WA, Walker A, Rushmer R, Elder L, and Methven H. Pain assessment with interactive computer animation. Pain 1993;53:347-51.
5. Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB. Discordance of databases designed for claims payment versus clinical informations systems. Ann Intern Med 1993;119:844-50.
6. Tu SW, Kemper CA, Lane NM, Carlson RW, and Musen MA. A methodology for determining patients' eligibility for clinical trials. Meth Inform Med 1993;32:317-25.

Address of the author:
George Hripcsak, M.D.,
Center for Medical Informatics,
Columbia-Presbyterian Medical Center,
161 Fort Washington Avenue,
New York, NY 10032,
USA