MEDINFO 2001
V. Patel et al. (Eds)
Amsterdam: IOS Press

An Automated Tool for an Analysis of Compliance to Evidence-based Clinical Guidelines

Brent A. Metfessel[a]

[a]*HealthTechnics, Inc., Eagan, Minnesota, USA*

Abstract

Evidence-based clinical guidelines have been developed in an attempt to decrease practice variation and improve patient outcomes. Although a number of studies and a few commercial products have attempted to measure guideline compliance, there still exists a strong need for an automated product that can take as input large amounts of data and create systematic and detailed profiles of compliance to evidence-based guidelines. The Guideline Compliance Assessment Tool is a product presently under development in our group that will accept as input medical and pharmacy claims data and create a guideline compliance profile that assesses provider practice patterns as compared to evidence-based standards. The system components include an episode of care grouper to standardize classifications of illnesses, an evidence-based guideline knowledge base that potentially contains information on several hundred distinct conditions, a guideline compliance scoring system that emphasizes systematic guideline variance rather than random variances, and an advanced data warehouse that would allow drilling into specific areas of interest. As provider profiling begins to shift away from a primary emphasis on cost to an emphasis on quality, automated methods for measuring guideline compliance will become important in measuring provider performance and increasing guideline usage, consequently improving the standard of care and the potential for better patient outcomes.

Keywords:

Evidence-Based Medicine; Physicians' Practice Patterns; Guideline Adherence; Medical Informatics; Practice Guidelines

Introduction

Even for the same illness conditions, a wide variation in practice patterns exists across different geographical locations and clinical practices. For example, major age, gender, and inter-practice variations were found in statin use in England and Wales, which was not well explained by patient diagnosis [1]. In the USA, physicians' recommendations for patient activities in chronic low back pain were widely variable and often overly restrictive [2]. In Swedish primary care practices, an estimated 10% savings in clinical chemistry costs could be obtained through optimizing and standardizing lab test ordering behaviors [3]. Developing countries also show wide practice pattern variation, complicated even further by difficulties in access to the medical literature [4].

To deal with the problem of clinical practice variation and potential care outcome problems, evidence-based clinical guidelines were created in an attempt to better standardize care and provide "best practices" based on the scientific literature. The goal was to encourage providers to utilize treatments with clear support from valid research studies. Problems occurred, however, in the actual implementation of such guidelines. First, care providers needed to be convinced that the guidelines are based on reviews of the most recent well-controlled studies [5], preferably with a "grading" system for the quality of the evidence. Second, no clinical guideline accounts for all unique patient presentations that may legitimize variances from the guideline recommendations for those patients. As a result, the guideline may be dismissed as "cookbook" medicine. Third, evidence is emerging that "passive" implementation of evidence-based guidelines (e.g. through didactic lectures or posted reminders) is considerably less effective than "active" implementation, where the physicians take initiative in the process, including discussion groups, feedback to clinicians, and/or surveillance by a clinical resource manager [6,7,8]. To actively implement guidelines, then, significant resources must be invested in the effort. A critical link in the process is the systematic method of measuring compliance to such guidelines on a large scale, which is necessary in an active implementation process, especially on more global scales. This accountability is needed to encourage effective behavior change which is the basic underlying reason for guideline implementation projects.

There have been some efforts underway in the public sector and in industry to measure compliance to components of

guidelines, but these projects have often been based on guidelines developed by expert panels rather than on a graded review of the scientific literature. In the United States, the Health Employer Data and Information Set (HEDIS), put out by the National Committee for Quality Assurance (NCQA) has made large strides forward in providing the infrastructure for development of computerized measures of certain aspects of care quality. Such measures include screening rates for breast and cervical cancer, frequency of retinal exams for diabetics, use of beta-blockers after acute myocardial infarctions, and other measures. However, these measures only cover a basic set of conditions and do not for the most part track treatment over the course of time or the progression of an illness. The "Patterns Profiler" commercial product (McKessonHBOC, Inc., San Francisco, California, USA) measures compliance to internally-developed expert-panel-based guidelines in terms of inappropriate procedures, procedure overutilization, or unusual procedure intensity. Consequently, this product is mainly geared toward utilization management and cost containment rather than true quality assessment; for example, it would not answer the question of whether a recommended drug was prescribed for a particular illness condition. A new product is needed that uses scientific evidence-based guidelines for a wide range of disease conditions and comprehensively measures compliance using commonly available data.

The Information Reporting Group at Health Risk Management, Inc. (HRM) is presently developing the Guideline Compliance Assessment Tool (GCAT) that will offer a thorough analysis of care quality based on scientific evidence-based guidelines.

Methods and Components of the System

Clinical Evidence Summaries

The Institute for Healthcare Quality, a subsidiary of HRM, is a vendor of scientific evidence-based guidelines in use by Health Plans and self-insured employer groups. One product that is produced as a result of the guideline development process is the Clinical Evidence Summary. Each Clinical Evidence Summary consists of a written review of the relevant literature for a particular illness condition, including a grading of the evidence -- for example, a well-designed randomized double-blind controlled study is graded relatively highly, while evidence based on textbooks or case reports is graded low. The guidelines and Clinical Evidence summaries are developed initially using Registered Nurse researchers in consultation with sponsoring specialty physicians. After this phase of development is completed, panels made up of physician experts provide feedback on the guideline and the Clinical Evidence Summary, suggest any needed final modifications, and give final approval to the guideline. It is the Clinical Evidence Summaries that will be encoded in the GCAT Knowledge Base.

Episode Treatment Groups

To more comprehensively analyze evidence-based guideline compliance, one needs to be able to track the time of onset and resolution of an illness and the longitudinal progression of an illness over time. To perform this analysis, an external product known as Episode Treatment Groups (ETGs) (Symmetry Health Data Systems, Inc., Phoenix, Arizona, USA) will be used to classify medical and pharmacy claims data into episode of care groupings that are based on disease states (using ICD diagnosis codes). Use of this episode grouper makes temporal relationships between interventions a valid quality measure in addition to providing a natural case-mix adjustment mechanism which may alter the way a guideline is applied to the case (i.e. illness classes can be further broken down into severity categories). ETG groupings are then matched to the Clinical Evidence Summary conditions.

Evidence-Based Guideline Knowledge Base

The narrative recommendations from the Clinical Evidence Summaries will be encoded into into machine-readable (parsable) rules. Although initially a manual process, natural language processing algorithms will likely be able to automate this translation process in future versions. For such encoding, ICD codes will be used for diagnosis coding and Current Procedural Terminology (CPT) codes from the American Medical Association will be used for procedure coding. If a different procedural coding standard is used in a certain area, a mapping to CPT codes can readily be developed. For the knowledge base, such encodable rules include, but are not limited to:

- A procedure is commonly indicated for a specific condition and should be performed. An example of such a procedure would be regular glycosylated hemoglobin tests for diabetics. Performing this procedure at appropriate intervals would increase the compliance score for the episode.

- A procedure is not generally indicated or is contraindicated for the specific condition. Performance of this procedure within an episode of that condition would decrease the compliance score.

- A procedure is indicated only after a specified time interval from episode onset. For instance, uncomplicated low back pain is first treated conservatively for about 30 days. If the episode is not resolved at that point, a magnetic resonance imaging (MRI) scan can, and normally should, be performed, in which case the compliance score would increase.

- A procedure can be performed for the condition but is limited in the number of times it should be performed. For example, multiple MRI scans for a surgical herniated disk are generally not recommended and would decrease the compliance score.

- A procedure is indicated only after a specified time interval from episode onset, but the compliance score for the episode *decreases* if the procedure is not performed. For example, it is necessary to perform a colectomy after a patient has had ulcerative colitis for 10 years due to the increased cancer risk (assuming 10 years of data was available).

- A procedure is indicated only after another procedure is performed. An example of this would be the performance of a screening lab test prior to a more extensive diagnostic workup.

- Inpatient utilization measures and other setting of care metrics will also be analyzed. Inpatient admissions for surgical procedures generally performed outpatient, consistently long inpatient lengths of stay (exposing the patient to the risks of hospitalization such as nosocomial infections), or unnecessary use of assistant surgeons will have a negative impact on guideline compliance. Conversely, performing a procedure on an outpatient basis that is normally performed inpatient may also decrease compliance scores, since that could subject the patient to unnecessary risks.

- The profiling of pharmaceutical compliance to evidence-based guidelines is a unique feature of the system, and is not generally captured in utilization review products. If pharmacy claims are available, one can determine if a prescribed medication is appropriate for the illness class based on the encoded rules. One can also analyze whether a drug was prescribed in the proper sequence for the condition. For example, a patient with new-onset hypertension should not generally be treated with a third-line drug without using a first or second-line drug first.

Guideline Compliance Scoring System

A scoring system for evaluating the level of guideline compliance for a particular illness classification is under development that will take into account the following factors:

- The grade of evidence behind the rule used for that part of the score. The higher the grade of evidence in the original Clinical Evidence Summary, the greater the rule weight in the guideline compliance score.

- For any particular rule, differential scoring will be allowed based on full, partial, or complete non-compliance to a rule. For example, a glycosylated hemoglobin test should be performed at least every six months for diabetics. However, yearly tests will score better than no tests at all but not as well as tests every six months.

- The scores will emphasize systematic variances from a guideline (across many episodes of the illness condition) rather than occasional variances due to unique patient factors. Repeated variances can be shown in an "alerts" section of a report and be used for quality initiatives.

The guideline compliance scoring system will be based on the following factors, related to the structure of the Knowledge Base:

- Degree of performance of generally necessary procedures.

- Degree of performance of generally unnecessary or harmful procedures.

- Appropriateness of procedure frequency.

- Timing of procedure with respect to progression of illness.

- Timing of procedure with respect to prerequisite procedures.

- Appropriateness of care setting (e.g. inpatient vs. outpatient)

- Appropriateness of inpatient length of stay.

- Appropriateness of pharmaceutical utilization.

- Miscellaneous (e.g. assistant surgeon).

An overall score for the episode will be obtained, with subscores for each of the previous factors where applicable. The scores and subscores will be integers ranging from 0 to 100, with 0 being completely non-compliant and 100 being compliant in every way. The score can also be affected by how "systematic" or "random" the guideline variances are within providers or provider groups. Systematic variance (repeated variances on the same measure, such as repeated neglect of annual diabetic retinal exams) will adjust the score further downward, while apparently random variances (scattered variances on different measures) that seem likely due to unique patient factors will cause the score to compensate upward. This score adjustment will be done at the aggregate level, with the level of aggregation to be selected by the user.

Data Warehouse

The claims and administrative databases for input and the GCAT output will be stored in a data warehouse using a Relational On-Line Analytic Processing (ROLAP) model. This data warehouse will provide the basis for all reporting and analysis on the data.

Information Delivery

A Web-based reporting delivery system will be developed that allows a user to identify areas of interest and drill down on those specific areas. For example, in cases of low back pain one may find a provider with a low score on "timing of procedure with respect to progression of illness". A user can then select that measure and through drill down find that this provider systematically orders MRI scans at initial presentation. This information can then be used to support active guideline implementation initiatives.

Initial Development

The initial phase of the GCAT project will be based on the Clinical Evidence Summaries for five significant illnesses: Hypertension, depression, diabetes mellitus, congestive heart failure, and asthma / reactive airway disease. Subsequently, the number of covered illnesses will be greatly expanded. The Institute of Healthcare Quality has Clinical Evidence Summaries that cover over 300 illnesses at present.

The overall flow of the system is shown in Figure 1.

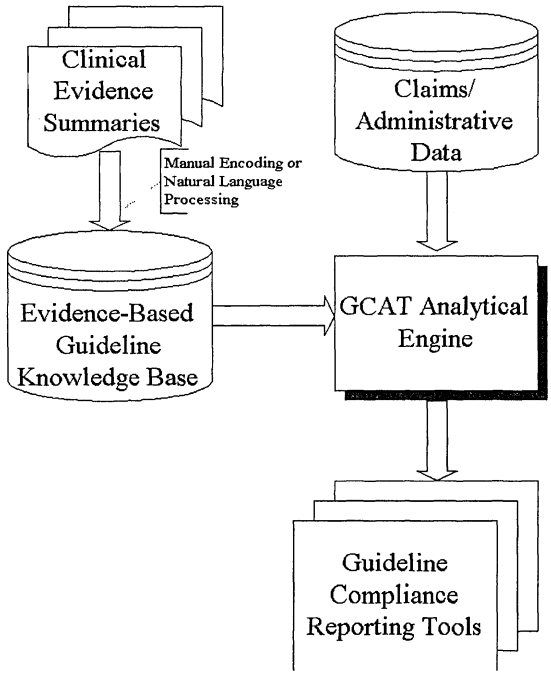

Figure 1 - Overall process flow for GCAT

Discussion

Potential Advantages of the Tool

Studies have repeatedly shown that well-designed implementation of evidence-based clinical guidelines can significantly improve practice patterns and appropriateness of drug prescribing behavior [9,10]. A system for automating guideline compliance assessment on a large scale can better integrate evidence-based guideline implementation initiatives with quality improvement, increasing the likelihood of visible, practical results. Some unique features and advantages of this tool include:

- The integration of the measurement of both cost-effectiveness and quality of care together to optimize the care process.
- The tool uses scientific evidence-based guidelines developed using rigorous evidence grading standards that encourage clinician buy-in to the process.
- The use of an episode of care grouper that enables tracking of the progression of the illness over time and brings forth temporal contingencies in treatment, creating opportunities for a more comprehensive assessment of guideline compliance.
- The ability to include the appropriateness of pharmaceutical interventions add an additional important dimension of quality analysis.
- A scoring algorithm that emphasizes systematic rather than random variance from the guidelines, thus allowing the "art" of medicine to be practiced. This will lead to greater provider acceptance of the system.

Uses of GCAT

The system has a number of potential practical uses:

- Consumers can use the results to select providers that are particularly competent in medical conditions of interest.
- Health plans and administrators can use the results for provider education, quality improvement, optimization of provider networks, and adoption of care standards.
- Providers can compare themselves to their peers and to guideline standards and note where self-education and improvement is needed.

Limitations of the System

Claims data is limited in its ability to depict the true process of medical care. For example, laboratory results, vital signs, progress notes, and history and physical exam

findings are not available using claims data. In addition, it may be difficult to "assign" episodes of care (and a guideline compliance score) to providers in episodes where many providers take part. Some options include: assigning the episode to the provider with the highest cost or number of visits, or "fragmenting" the episode between the different providers. Each method has its advantages and drawbacks [11], and the decision on which to use is individual.

Future Enhancements

Such enhancements include the expansion of the covered disease states, the use of natural language processing algorithms to make the conversion of the Clinical Evidence Summaries into machine-readable knowledge bases more efficient, and the inclusion of other data streams such as lab results and the automated medical record, which would greatly expand the universe of measurable quality factors.

Conclusion

The Guideline Compliance Assessment Tool has the potential to significantly advance the translation of scientific research into clinical practice through the use of large-scale automated methods to comprehensively track guideline compliance, thus assisting provider education and quality improvement initiatives in the global medical community.

Acknowledgements

I would like to thank Ken Buchanan (Vice President, Information Reporting) and the other members of the Information Reporting Group at HRM for their support and efforts in this project.

References

[1] Majeed A, Moser K, and Maxwell R. Age, sex, and practice variations in the use of statins in general practice in England and Wales. *J Public Health Med* 2000: 22(3) pp. 275-9.

[2] Rainville J, Carlson N, Polatin P, and Gatchel RJ. Exploration of physicians' recommendations for activities in chronic low back pain. *Spine* 2000: 25(17) pp. 2210-20.

[3] Larsson A, Palmer M, Hulten G, and Tryding, N. Large differences in laboratory utilisation between hospitals in Sweden. *Clin Chem Lab Med* 2000: 38(5) pp. 383-9.

[4] Page J, Heller RF, Kinlay S, Lim LL, and Qian W. Where do developing world clinicians obtain evidence for practice: a case study on pneumonia. *J Clin Epidemiol* 2000: 53(7) pp. 669-75.

[5] Crim C. Clinical practice guidelines vs actual clinical practice: the asthma paradigm. *Chest* 2000: 118 (2 Suppl) piii: 62S-64S.

[6] Frankel HL, Fitzpatrick MK, Gaskell S, and Hoff WS. Strategies to improve compliance with evidence-based clinical management guidelines. *J Am Coll Surg* 1999: 189(6), pp. 533-8.

[7] Bergman DA. Evidence-based guidelines and critical pathways for quality improvement. *Pediatrics* 1999: 103 (1 Suppl E) piii. 225-32.

[8] Onion CW, and Bartzokas CA. Changing attitudes to infection management in primary care: a controlled trial of active versus passive guideline implementation strategies. *Fam Pract* 1998: 15(2) pp. 99-104.

[9] Bell CM, Ma M, Campbell S, Basnett I, Pollock A, and Taylor I. Methodological issues in the use of guidelines and audit to improve clinical effectiveness in breast cancer in one United Kindgom health region. *Eur J Surg Oncol* 2000: 26(2) pp. 130-6.

[10]Alexander KP, Peterson ED, Granger CB, Casas AC, Van de Werf F, Armstrong PW, Guerci A, Topol EJ, and Califf RM. Potential impact of evidence-based medicine in acute coronary syndromes: insights from GUSTO-Iib. Global Use of Strategies to Open Occluded Arteries in Acute Coronary Syndrmes trial. *J Am Coll Cardiol* 1998: 32(7) pp. 2023-30.

[11]Metfessel BA. Specialist profiling using claims and administrative databases. *AJIH* 1998: 1(4) pp. 168-174.

Address for Correspondence

Brent A. Metfessel, MD, MS
Director of Medical Informatics
Health Risk Management, Inc.
10900 Hampshire Ave South
Minneapolis, MN 55438-2306
Ph: 1-800-824-3882 ext. 3542
Fax: 1-952-829-3578
Email: bmetfess@hrmi.com