

A comprehensive appropriateness of prescribing questionnaire was validated by nominal consensus group

Anne M. Holbrook^{a,b,c,d,*}, Vladan Janjusevic^a, Charlie H. Goldsmith^{a,d},
Ivan Y. Shcherbatykh^{a,d,2}
for the COMPETE Investigators

^aCentre for Evaluation of Medicines, St. Joseph's Healthcare, McMaster University, 105 Main Street East, P1, Hamilton, Ontario, Canada L8N 1G6

^bDivision of Clinical Pharmacology & Therapeutics, McMaster University, 105 Main Street East, P1, Hamilton, Ontario, Canada L8N 1G6

^cDepartment of Medicine, McMaster University, 105 Main Street East, P1, Hamilton, Ontario, Canada L8N 1G6

^dDepartment of Clinical Epidemiology and Biostatistics, McMaster University, 105 Main Street East, P1, Hamilton, Ontario, Canada L8N 1G6

Accepted 15 January 2007

Abstract

Objective: To develop and validate a comprehensive Appropriateness of Prescribing Evaluation Questionnaire (APEQ) suitable for human and computer use.

Study Design and Setting: This study was part of an ongoing research program examining the effectiveness and cost-effectiveness of computerized prescribing decision support for providers, patients, and drug policy. A nominal group consensus process involved physicians, both primary care physicians and specialists, pharmacists, drug plan managers, patients, patient advocates, and pharmaceutical industry. Structured case scenarios of musculoskeletal problems were used to evaluate APEQ's validity and responsiveness.

Results: Seventeen panelists evaluated 72 patient scenarios in two rounds. Their ratings of appropriateness, assessed by ANOVA, showed significant agreement with the experts' scores in the two rounds, which evaluated appropriateness and responsiveness, respectively. Interrater and intrarater agreement was moderate to good.

Conclusion: This formal assessment suggests that APEQ has reasonable validity, reliability, and responsiveness. Such tools could be very useful in e-prescribing and e-claims reimbursement environments and should be further explored. © 2007 Elsevier Inc. All rights reserved.

Keywords: Questionnaire; Validation; Reliability; Prescribing; Appropriateness; Computerized drug utilization evaluation

1. Background

Expenditures for pharmaceuticals continue to consume proportionately more health care resources leading to an unprecedented demand for evidence of value for money [1]. "Value" implies tangible benefit to patients in terms of quality or quantity of life, or to the system in terms of decreased demand for resources or improved efficiency of services. Improving the quality of prescribing relies on a combination of accurate patient assessment, detailed knowledge of current evidence on therapeutic options

including their cost-effectiveness and limitations, and the ability to negotiate the best advice with the patient's values and expectations. The measurement of appropriateness of prescribing has largely been left to implicit judgment by "experts" [2,3]. Based on consensus opinion, a number of studies have concluded that the prevalence of inappropriate drug prescribing is high, especially among elderly people [4–6]. However, these judgments have often proven to be flawed because of lack of explicit consideration of evidence [7,8] or lack of consideration of key details of the clinical situation [9,10]. For example, the lack of consideration of a previous intolerance or lack of response to a first-line medicine may lead to an erroneous conclusion of inappropriateness. Other key details include diagnoses, medical history, allergies, and laboratory results, all of which require validation. Even when an evaluation of drug appropriateness has been complete and thorough, expert opinion is subject to individual bias

¹ Dr. Holbrook is the recipient of a Career Investigator award from the Canadian Institutes of Health Research.

² Dr. Shcherbatykh is the recipient of a Father Sean O'Sullivan Studentship Award.

* Corresponding author. Tel.: 905-522-1155x5269; fax: 905-528-7386.
E-mail address: holbrook@mcmaster.ca (A.M. Holbrook).

What is new?

1. Evidence-based, comprehensive appropriateness of prescribing rating tools that have algorithms sufficiently explicit to be computerized are rare.
2. We developed a comprehensive appropriateness of prescribing rating template (APEQ) and then a drug-specific example (MAPEQ) that could be computerized.
3. A formal nominal consensus validation process confirmed the reliability and validity of this comprehensive appropriateness of prescribing rating tool.

and cannot easily be applied or replicated across a wide variety of patients.

Computerized drug utilization evaluation (eDUE) has been embraced across Canada and internationally, primarily as an efficient way to influence prescribing through automated reimbursement rules. However, as currently carried out, eDUE uses administrative data, which have severe limitations due to the lack of most of the key details outlined above [11–14]. Electronic medical records that have structured charting entry, detailed longitudinal patient data, and organization of the data into discrete fields for analysis are an innovation with great potential as a platform for high-quality eDUE and targeted, patient-specific interventions to influence prescribing [12,15]. An electronic appropriateness of prescribing rating scale that evaluates prescriptions as they are being prescribed or claims as they are being processed, with the additional information required to make an expert evaluation, would be extremely useful both to streamline intelligent adjudication of claims and to intervene to suggest alternatives to the prescriber when necessary. Not only would such a tool advance current eDUE, which monitors more than 6 million claims daily (approximately \$120 billion worth annually) in Canada and the United States [16,17], but it would form the necessary foundation for real-time intervention at the point of care. Furthermore, the uptake of e-prescribing and electronic order entry depends on intelligent, patient-specific advice. The translation of “expert judgment” into explicit algorithms detailed enough to be programmable has not been accomplished for a comprehensive, high-quality appropriateness of prescribing tool. Explicit rules and algorithms also have the advantage of allowing for reliability and validity testing, an important methodologic confirmation of quality.

A search of MEDLINE and EMBASE (1950 to December 2006) revealed two published questionnaires designed to comprehensively rate the quality of a prescription [18–20]. However, the Medication Appropriateness Index

and Prescribing Appropriateness Index described do not emphasize the use of high-quality evidence, have limited generalizability (because they are meant to be applied by clinical pharmacists rounding on hospital or nursing home wards), and use implicit rather than explicit judgment to score each of their questions. The implicitness makes them unsuitable for electronic use. We therefore set out to develop and validate the Appropriateness of Prescribing Evaluation Questionnaire (APEQ) as an explicit, evidence-based, programmable, patient-specific tool to measure the appropriateness of prescriptions.

2. Methods

A core team of five experienced clinical pharmacologists and pharmacists identified 10 different domains of appropriateness for a prescription. A series of prescriptions were used to determine that each of the 10 items was suitably independent from each other and was recognizable as a criterion. The criteria related to the following: indication based on the diagnostic labels applied by the physician, indication based on the patient’s previous therapy, optimal drug choice in terms of benefit/risk evidence compared to the drug’s peers, dose, duration of treatment, directions for use, duplication of medications prescribed, contraindications related to comorbidity or concomitant medications, and comparative cost-effectiveness. For each of these criteria, a score of 0 (totally inappropriate), 1 (partially appropriate), 2 (totally appropriate), or DK (don’t know, meaning insufficient information available) could be assigned.

To apply and validate the APEQ approach, a disease-specific, Musculoskeletal Appropriateness of Prescribing Evaluation Questionnaire (MAPEQ) for nonsteroidal anti-inflammatory drugs (NSAIDs) was developed. This involved a systematic review of the literature for each criterion, followed by development of a scoring grid, carried out in consultation with a provincial musculoskeletal guidelines expert panel [21], the coinvestigators, and the core panel. At the completion of the systematic review and consultation, it became clear that there was insufficient evidence to score two of the 10 criteria—directions for taking the medication and optimal drug choice based on benefit/risk evidence. For the former, no good evidence could be found to support the common nostrum to take NSAIDs with food. For the latter, although expert opinion provided many and often conflicting recommendations, no high-quality evidence supported differences among NSAIDs based on benefit/risk ratios.

The eight remaining criteria in MAPEQ were clustered into three domains for analytic feasibility: Domain 1 included indication based on the diagnosis, indication based on the prior therapy, and duplication of medication, and cost. Domain 2 included daily dosage prescribed and duration of treatment. Domain 3 included clinically important drug–drug interactions and clinically important drug–disease contraindications. Scoring of each domain was

collapsed into three alternatives—totally appropriate (score = 2), partially appropriate (score = 1), and totally inappropriate (score = 0). With three possible scores per domain, there were 27 (3 × 3 × 3) possible combinations of scores. Total MAPEQ score could vary from 0 to 6 (three domains, each with a possible score of 0–2).

In preparation for the validation consensus meeting, realistic primary care patient scenarios were created to represent a variety of musculoskeletal disorders as well as the full gradient of prescribing quality scoring, from totally inappropriate (0) to totally appropriate (6). A sample case is shown in Fig. 1. Intrarater reliability testing was built in by altering minor details for nine of the cases and re-presenting them. Thus, a total of 36 (27 plus nine) patient cases were constructed. Each of the 36 patient scenarios was scored by two MAPEQ scoring experts (a senior clinical pharmacologist and a methodologist, both involved in MAPEQ development), whose scores were labeled “expert scores.” Because we also wished to evaluate the responsiveness of MAPEQ, a second set of 36 scenarios was constructed in which the original 36 patient cases were presented again as the baseline visit and, for each case, information on a 6-month follow-up visit was added. A sample follow-up case is presented in Fig. 2. In this set, the scoring outcome was change in prescribing quality—better, worse, or no change. These

follow-up visit scenarios were constructed to represent a breadth of scores from much worse to much better.

The APEQ validation exercise used a nominal group technique [22–24]. This process consisted of two rounds of a face-to-face daylong meeting during which panel members rated, discussed, and rated again each case. The consensus panelists were chosen to represent all stakeholders in the domain of appropriateness of prescribing: medical specialists, family physicians, clinical pharmacologists, pharmacists, drug programs administrators, the pharmaceutical industry, and consumers. A total of 17 panelists were recruited and allocated to two groups, each group with a representative of each stakeholder group. Two facilitators, experts in nominal group process, conducted the meeting. The panelists were told that they were rating appropriateness of prescribing, but were given no details about rating scales, the scoring criteria, domains, or repeat cases. The results of the appropriateness of prescribing session (first session) were to confirm the face and construct validity of the MAPEQ tool by comparing the results of panelists with scores assigned by the expert raters. Each panelist individually rated prescribing for each scenario as appropriate, inappropriate, or indeterminate. After this first round of scoring was completed, each group’s facilitator summarized responses for each case. Agreement was defined as

| Scenario 60A | | | |
|--|------------------------------|---|--------------------------|
| Female, 53 y/o - New diagnosis of <i>rheumatoid arthritis</i> . 3 month history morning stiffness (1.5 hours duration), decreased appetite, fatigue, generalized muscle/ joint pain. Bilateral, symmetrical swelling, tenderness and warmth of MCP and PIP joints of hands and MTP joints of feet. | | | |
| Current Med. Problems | Duration | Current Medications | Duration |
| Menopause/hot flashes | 1 y | Estrogen 0.625 mg/d Medroxyprogesterone 2.5 mg/d Vit C Vit E | 1 y 1 y 3 y 3 y |
| Past Med. History | Start/Duration | Past Meds | Start/Duration |
| Nil | | | |
| Labs. | Ref. Ranges | Prescribed Therapy: Ibuprofen 600 mg tid 1 m supply, 5 repeats. Rest, ice, compression, exercise. | |
| RF pos | | | |
| ESR 52 | < 30 mm/h (F) | | |
| Cr 52 | 50-110 umol/L | | |
| ANA neg | | | |
| Hgb 106 | 110 – 180 g/L (F) | | |
| Plt 480 | 150-450 x 10 ⁹ /L | | |

APPROPRIATENESS (please check only one box)

| | | |
|--------------------------|--------------------------|--------------------------|
| APPROPRIATE | NOT APPROPRIATE | CAN'T TELL |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

AFTER DISCUSSION, DO YOU WANT TO RE-EVALUATE?

| | | |
|--------------------------|--------------------------|--------------------------|
| APPROPRIATE | NOT APPROPRIATE | CAN'T TELL |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Fig. 1. Sample case: evaluating appropriateness.

| Scenario 60B | | SIX MONTHS LATER | |
|--|------------------------------|---|----------------|
| Female, 53 y/o. - <i>RA</i> – Warm, swollen, tender joints: PIP, MTP, ankles over last 3 days. Recently started Piroxicam but hasn't taken for last 3 days. Morning stiffness 1.5 h – relieved by hot shower. Diagnosed with NSAID – ulcer in walk-in clinic 3 days ago. | | | |
| Current Med. Problems | Duration | Current Medications | Duration |
| RA | 6 m | Piroxicam 20 mg po daily | 1 w |
| NSAID-induced ulcer | 3 d ago | Omeprazole 20 mg po bid | 3 d |
| HTN | 3 m | Atenolol 50 mg/d | 3 m |
| Menopause/hot flashes | 1.5 y | | |
| Past Med. History | Start/Duration | Past Meds | Start/Duration |
| | | Ibuprofen 600 mg po tid | 6 m ago X 5 m |
| Labs. | Ref. ranges | Prescribed Therapy: D/C Piroxicam Naproxen 500 mg po bid 1msupply, 10repeats. Aerobic pool, rheumatology referral. | |
| BP 138/82 | | | |
| ESR 41 | < 30 mm/h (F) | | |
| Hgb 114 (2m ago) | 110 – 180 g/L (F) | | |
| Cr 82 | 50-110 umol/L | | |
| Alb 37 | 35-50 g/L | | |
| Plt 295 (2 m ago) | 150-450 x 10 ⁹ /L | | |

The appropriateness of prescribing for the **SCENARIO SIX MONTHS LATER...**

| | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|
| IMPROVED | WORSE | NO CHANGE | CAN'T TELL |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

AFTER DISCUSSION, DO YOU WANT TO RE-EVALUATE?

| | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|
| IMPROVED | WORSE | NO CHANGE | CAN'T TELL |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Fig. 2. Sample case in follow-up: evaluating responsiveness.

at least 70% of participants applying the same rating (appropriate, inappropriate, or indeterminate) for the case. Cases for which agreement was not reached were then the topic of discussion for as much time as panelists needed. The facilitators ensured that each panelist had a chance to discuss his or her views. After the discussion, panelists were allowed to reconsider their ratings for any case, and this second round produced the final scores. Interrater reliability was calculated based on these final ratings.

In the second session, each of the 36 cases used in the morning session was presented again, this time with a 6-month “follow-up note” that represented the patient’s subsequent visit. Panelists rated each of these case summaries as improved, no change, worse, or indeterminate. Again, once the scoring was completed, discussion was encouraged for cases for which agreement was not reached following which the panel members had the opportunity to score these cases again. The intent was to evaluate the responsiveness of MAPEQ by examining whether case scenarios illustrating a clinically important change according to the expert scores yielded similar patterns when scored by the panel.

The overall results that determined the validity of the MAPEQ were analyzed using ANOVA examining the main effects and two-factor interactions. Associations between MAPEQ criterion (expert) scores and panelists’ scoring for responsiveness were analyzed with Pearson’s correlation and the Spearman rank correlation coefficient. Interrater reliability was estimated using an intraclass correlation coefficient (ICC) with a general linear model procedure and two-way random effects design with interactions; intrarater reliability was estimated using kappa. All of the statistical procedures were performed with SPSS, version 10.0 (SPSS, Chicago, IL, USA).

3. Results

The analysis was based on 1,224 ratings (17 raters, 36 patient scenarios, two rounds). No problems were

encountered with the nominal group process, either with the discussion, the timelines, or the clarity of process.

3.1. MAPEQ validation—appropriateness

Table 1 shows the mean proportion of 36 patient scenarios that were scored “appropriate” by the panelists, compared to the expert scores. Despite having no knowledge of the scoring system, panelists’ ratings of the patient scenarios generally correlated with the expert scores. The results of the ANOVA of the panelist ratings compared to the expert scores show a statistically significant correlation (Table 1). In addition, there is a highly significant interaction between Domain 1 and Domain 2, demonstrating that the panel members took indications, duplication, and cost into consideration when rating dosing and duration and vice versa. Domain 3 remained independent, and the other interactions were not statistically significant at the 5% level.

Table 2 shows the mean proportion of scenarios scored as “appropriate” for total MAPEQ score.

The higher the expert scores, the more frequently the prescribing was viewed as appropriate by the panelists. However, the panelists were more critical in their ratings as relatively high expert scores were required before the prescribing was judged as appropriate by the panelists. The ANOVA ($F = 6.65$, $P < 0.001$) confirmed that the pattern of panelist scoring was significantly similar to that of the experts’ scoring. Correlation between MAPEQ total score and the proportion judged “appropriate” by the panelists was further analyzed. Pearson’s correlation coefficient was 0.53 ($P < 0.001$), showing moderate positive association.

3.2. MAPEQ validation—responsiveness

The results from the responsiveness session assessed whether the panelists’ ratings agreed with the MAPEQ expert scores of predesigned changes in each scenario. Scenario changes were designed to represent the full spectrum of changes in appropriateness of prescribing from

Table 1
Panelist versus expert scores for appropriateness domains

| MAPEQ expert score ^a | Domain 1 | | Domain 2 | | Domain 3 | |
|---------------------------------|-----------------------------|---|-----------------------------|---|-----------------------------|---|
| | Number of patient scenarios | Mean proportion judged appropriate by panelists | Number of patient scenarios | Mean proportion judged appropriate by panelists | Number of patient scenarios | Mean proportion judged appropriate by panelists |
| 0 | 10 | 0.05 | 14 | 0.06 | 11 | 0.03 |
| 1 | 11 | 0.05 | 11 | 0.04 | 12 | 0.10 |
| 2 | 15 | 0.12 | 11 | 0.15 | 13 | 0.11 |

Notes: Domain 1 = diagnosis and therapy indications, duplication, cost; ANOVA F -value = 5.85, P -value = 0.008.

Domain 2 = dosing, duration; ANOVA F -value = 5.36, P -value = 0.012.

Domain 3 = drug interactions, contraindications; ANOVA F -value = 3.60, P -value = 0.042.

Interaction between Domain 1 and Domain 2: ANOVA F -value = 8.88, P -value = < 0.001.

^a 0 = totally inappropriate, 1 = partially appropriate, 2 = totally appropriate.

Table 2
Panelist versus expert scores for total MAPEQ score

| Total MAPEQ score—expert | Number of patient scenarios | Proportion judged appropriate by panelists |
|--------------------------|-----------------------------|--|
| 0 | 2 | 0.04 |
| 1 | 3 | 0.05 |
| 2 | 8 | 0.04 |
| 3 | 7 | 0.06 |
| 4 | 10 | 0.06 |
| 5 | 4 | 0.17 |
| 6 | 2 | 0.36 |

much better to much worse. Table 3 shows the pattern of scoring of change by the panelists.

In general, the panelists' scoring patterns did follow the predesigned changes of MAPEQ total scores. ANOVAs for each panelist's scores are presented in Table 4.

These show a highly significant association between panelist ratings of change and the expert scores, best described by a straight line (slope). For expert scores of "no change," the panelists still were detecting change, which may represent a type of expectation bias. Pearson's correlation coefficients for proportion of cases with improved appropriateness scores (0.455, $P = 0.005$) and proportion with worse scores (-0.682 , $P < 0.0001$) showed highly significant correlations with the expert scores.

3.3. MAPEQ reliability: interrater and intrarater

The interrater reliability of the panelists' scoring for the appropriateness of prescribing session was estimated using ICC. The ICC was 0.62 (95% confidence interval [CI]: 0.55, 0.70), representing moderate to substantial agreement [25]. The intrarater reliability of the panelist scoring for both sessions was estimated using kappa (κ). For the scoring of appropriateness, the estimated $\kappa = 0.58$ (95% CI: 0.45, 0.72). For the scoring of responsiveness (rating of change in appropriateness), the estimated $\kappa = 0.42$ (95% CI: 0.30, 0.53). These estimated kappa values represent moderate agreement.

Table 3
Panelist versus expert scores for MAPEQ responsiveness

| Actual change in MAPEQ total score ^a | Number of patient scenarios | Mean proportion "improved" as judged by panelists | Mean proportion "no change/cannot tell" as judged by panelists | Mean proportion "worse" as judged by panelists |
|---|-----------------------------|---|--|--|
| -3 | 6 | 0.02 | 0.29 | 0.70 |
| -2 | 5 | 0.23 | 0.45 | 0.33 |
| -1 | 5 | 0.06 | 0.32 | 0.63 |
| 0 | 2 | 0.10 | 0.47 | 0.44 |
| 1 | 4 | 0.27 | 0.58 | 0.16 |
| 2 | 5 | 0.34 | 0.39 | 0.28 |
| 3 | 6 | 0.27 | 0.58 | 0.16 |
| 4 | 3 | 0.40 | 0.52 | 0.10 |
| Overall | 36 | 0.204 | 0.439 | 0.367 |

^a As assessed by expert scorers.

4. Discussion

A questionnaire for application in clinical settings should be valid, reliable, and able to detect clinically important changes [26–28]. The main goal of this study was to validate our comprehensive prescribing evaluation tool, utilizing the disease-specific MAPEQ as an example. The results of the validation exercise demonstrated adequate face and content validity of the MAPEQ. Indeed, although the panelists were not aware of our appropriateness criteria or analytic algorithm, they implicitly applied the concept of domains while scoring the appropriateness of prescribing. Furthermore, the scoring of the panelists was in general agreement with the experts' ratings, thus providing acceptable construct validity. All three analytic domains were important to the overall score. Domains 1 and 2 had significant interactions, whereas Domain 3 was perceived independently. The comparisons of panelist ratings with the total MAPEQ score provided further evidence of the instrument's validity.

The evaluation of responsiveness demonstrated the good ability of MAPEQ to capture the underlying change in prescribing, varying from substantial deterioration to important improvement. The correct assessment of patients whose condition remained unchanged proved to be a more difficult task probably due to expectations for change by the panel members. Finally, some of the analyses of panelists' reliability showed only moderate agreement, a common finding in the clinical reasoning literature [29–31].

This study has several limitations. Although the nominal group technique represents a well-recognized qualitative methodology for generation, prioritization, and evaluation of ideas [24,32], it can only capture the thoughts of the experts involved in the group decision-making process. As well, although the patient scenarios utilized in the validation exercise presented a variety of musculoskeletal disorders and a broad spectrum of quality scores, and were clinically realistic, their representation of the entire spectrum of appropriateness of prescribing would be difficult to prove. A larger sample of cases might well have

Table 4
ANOVA of panelist versus expert scores for MAPEQ responsiveness

| | Degree of freedom | F-value | P-value |
|--|-------------------|---------|---------|
| For proportion “improved” | | | |
| All | 7 | 1.79 | 0.1292 |
| Slope | 1 | 8.37 | 0.0073 |
| Other | 6 | 0.69 | 0.6578 |
| For proportion “no change/cannot tell” | | | |
| All | 7 | 1.94 | 0.1005 |
| Slope | 1 | 7.07 | 0.0128 |
| Other | 6 | 1.09 | 0.3936 |
| For proportion “worse” | | | |
| All | 7 | 7.88 | <0.0001 |
| Slope | 1 | 38.73 | <0.0001 |
| Other | 6 | 2.74 | 0.0318 |

improved the representativeness or comprehensiveness of the cases, but it would have hampered the feasibility of the validation exercise in a practical sense due to increased complexity and time constraints. Finally, to obtain the evidence of APEQ’s general robustness, it should be formally tested for other important disease and therapeutics areas, such as diabetes or cardiovascular pathologies. This would require other disease-specific or drug-specific questionnaires.

Although expert judgment of the quality of prescribing may be subjective, nonexplicit, and expensive in terms of the personnel required, alternatives such as computerized appropriateness of prescribing evaluation tools remain a rarity. There are several good reasons for this. First, sophisticated computerized decision support as an emerging health technology is still in its infancy in terms of programming, use in practice, and evaluation of impact [33,34].

Second, the explicit algorithms required for computers often cannot cope with the multiple exceptions and clinical variations required to make them clinically applicable, nor can their creators cope with the constant vigilance required to keep them up to date. Third, attempts to remain faithful to the evidence supporting prescribing are always frustrated by the lack of evidence for key domains. For example, although a plethora of advice on directions for how to take medications is available through package inserts and pharmacies, almost none of this is based on high-quality (or moderate-quality) evidence. Fourth, because appropriateness is often disease specific and always medication specific, multiple APEQs would be required to evaluate most patient drug–disease situations followed by, perhaps, a “meta-APEQ” to handle the overall patient profile. Drug interactions, for example, are almost always reported as a duad—one drug interacts with another, rather than interactions among multiple drugs in the patient’s profile. This development and testing process is laborious and time consuming. Fifth, anchoring a medication on a diagnosis (indication by diagnosis) is often problematic. Much of medicine, in primary care in particular, is about shaping symptoms and signs into diagnoses. Patients are often

treated, particularly with symptomatic therapies, before a firm diagnosis is made. Furthermore, many conditions such as “arthritis” or “heart failure” are not specific diagnoses but syndromes and require further clarification in patient work-up to properly evaluate appropriateness of prescribing.

5. Conclusions

The APEQ, as illustrated by the MAPEQ, represents a reliable, responsive, and valid instrument that provides a comprehensive, evidence-based, and explicit measure of prescribing appropriateness. The fact that it is sufficiently explicit to be automated makes it unique and worthy of further investigation in electronic clinical decision support or automated reimbursement review environments.

Acknowledgments

The authors wish to acknowledge the expertise, industry, and collegiality of their valued colleague Dr. Janjusevic, who is now deceased. They also acknowledge the assistance with early versions of the MAPEQ by Dr. Laura Park-Wyllie.

The study was funded through the Ontario Program for Optimal Therapeutics by the Ontario Ministry of Health and Long-term Care.

References

- [1] Drug pricing: a comparison between Canada and other countries. An address by the Chairperson of the PMPRB—at the Institute for Research on Public Policy Conference—Toward a National Strategy on Drug Insurance: Challenges and Priorities, September 23, 2002, Toronto, Ontario. Available at: <http://www.pmprb-cepmc.gc.ca/cmfiles/sp02-irpp-e14NLB-482003-8592.pdf> 2002. (Accessed march 3, 2007).
- [2] Walton RT, Gierl C, Yudkin P, Mistry H, Vessey MP, Fox J. Evaluation of computer support for prescribing (CAPSULE) using simulated cases. *BMJ* 1997;315:791–5.
- [3] Beers MH, Ouslander JG, Rollingher I, Reuben DB, Brooks J, Beck JC. Explicit criteria for determining inappropriate medication use in nursing home residents. UCLA Division of Geriatric Medicine. *Arch Intern Med* 1991;151:1825–32.
- [4] Aparasu RR, Fliginger SE. Inappropriate medication prescribing for the elderly by office-based physicians. *Ann Pharmacother* 1997;31:823–9.
- [5] Hanlon JT, Schmader KE, Boulton C, Artz MB, Gross CR, Fillenbaum GG, et al. Use of inappropriate prescription drugs by older people. *J Am Geriatr Soc* 2002;50:26–34.
- [6] Tamblyn R. Medication use in seniors: challenges and solutions. *Therapie* 1996;51:269–82.
- [7] McLeod PJ, Huang AR, Tamblyn RM, Gayton DC. Defining inappropriate practices in prescribing for elderly people: a national consensus panel. *CMAJ* 1997;156:385–91.
- [8] Tierney WM, Overhage JM, McDonald CJ. Toward electronic medical records that improve care. *Ann Intern Med* 1995;122:725–6.
- [9] Soumerai SB, Lipton HL. Computer-based drug-utilization review—risk, benefit, or boondoggle? *N Engl J Med* 1995;332:1641–5.

- [10] Lipton HL, Bird JA. Drug utilization review in ambulatory settings: state of the science and directions for outcomes research. *Med Care* 1993;31:1069–82.
- [11] Drug utilization review: mechanisms to improve its effectiveness and broaden its scope. The U.S. Pharmacopeia Drug Utilization Review Advisory Panel. *J Am Pharm Assoc (Wash)* 2000;40:538–45.
- [12] Holbrook AM. Evaluation of data sources to support pharmaco-surveillance. Available at: http://www.hc-sc.gc.ca/sr-sr/finance/hprp-prpms/final/2004-pharmaco-data_e.html. Accessed January 2, 2007.
- [13] Hennessy S, Bilker WB, Weber A, Strom BL. Descriptive analyses of the integrity of a US Medicaid claims database. *Pharmacoepidemiol Drug Saf* 2003;12:103–11.
- [14] van Eijk ME, Krist LF, Avorn J, Porsius A, de Boer A. Do the research goal and databases match? A checklist for a systematic approach. *Health Policy* 2001;58:263–74.
- [15] Purves IN. The paperless general practice. *BMJ* 1996;312:1112–3.
- [16] IMS Health. Canada Rx Report 2005. Available at: www.imshealth-canada.com. Accessed January 23, 2006.
- [17] National Association of Chain Drug Stores. In 2001 retail pharmacies filled over 3 billion prescriptions totaling \$164 billion. Available at: www.nacds.org 2002. Accessed September 28, 2005.
- [18] Hanlon JT, Schmader KE, Samsa GP, Weinberger M, Uttech KM, Lewis IK, et al. A method for assessing drug therapy appropriateness. *J Clin Epidemiol* 1992;45:1045–51.
- [19] Samsa GP, Hanlon JT, Schmader KE, Weinberger M, Clipp EC, Uttech KM, et al. A summated score for the medication appropriateness index: development and assessment of clinimetric properties including content validity. *J Clin Epidemiol* 1994;47:891–6.
- [20] Barber N, Bradley C, Barry C, Stevenson F, Britten N, Jenkins L. Measuring the appropriateness of prescribing in primary care: are current measures complete? *J Clin Pharm Ther* 2005;30:533–9.
- [21] Holbrook AM for Ontario Musculoskeletal Therapy Review Panel. Ontario treatment guidelines for osteoarthritis, rheumatoid, arthritis, and acute musculoskeletal injury. 1st edition. Toronto: Queen's Printer of Ontario; 2000.
- [22] Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995;311:376–80.
- [23] Goldsmith CH, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. *J Rheumatol* 1993;20:561–5.
- [24] Delbecq AL, Van de Ven AH, Gustafson DH. Group techniques for program planning: a guide to nominal group and delphi processes. Middleton, WI: Green Briar Press; 1986.
- [25] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- [26] Streiner DL, Norman GR. Reliability. In: Streiner DL, Norman GR, editors. Health measurement scales. 2nd edition. New York: Oxford University Press; 1995. p. 105–27.
- [27] Juniper EF, Guyatt GH, Cox FM, Ferrie PJ, King DR. Development and validation of the Mini Asthma Quality of Life Questionnaire. *Eur Respir J* 1999;14:32–8.
- [28] Jackowski D, Guyatt G. A guide to health measurement. *Clin Orthop Relat Res* 2003;413:80–9.
- [29] Hatala R, Smieja M, Kane SL, Cook DJ, Meade MO, Nishikawa J. An evidence-based approach to the clinical examination. *J Gen Intern Med* 1997;12:182–7.
- [30] Koran LM. The reliability of clinical methods, data and judgments (second of two parts). *N Engl J Med* 1975;293:695–701.
- [31] Koran LM. The reliability of clinical methods, data and judgments (first of two parts). *N Engl J Med* 1975;293:642–6.
- [32] Cantrill JA, Sibbald B, Buetow S. Indicators of the appropriateness of long-term prescribing in general practice in the United Kingdom: consensus development, face and content validity, feasibility, and reliability. *Qual Health Care* 1998;7:130–5.
- [33] Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005;293:1223–38.
- [34] Holbrook A. Rule-based computerized reminders increased ordering of preventive services in an inpatient setting. *ACP J Club* 2002;136:74.