BMJ Open

Answering medical questions at the point of care: a cross-sectional study comparing rapid decisions based on PubMed and Epistemonikos searches with evidence-based recommendations developed with the GRADE approach

Ariel Izcovich,[1] Juan Martín Criniti,[1] Federico Popoff,[1] Martín Alberto Ragusa,[1] Cristel Gigler,[1] Carlos Gonzalez Malla,[1] Manuela Clavijo,[1] Matias Manzotti,[1] Martín Diaz,[1] Hugo Norberto Catalano,[1] Ignacio Neumann,[2] Gordon Guyatt[3]

[1]Department of Internal Medicine, Hospital Alemán, Buenos Aires, Argentina
[2]Department of Internal Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile
[3]Department of Evidence and Impact, Health Research Methods,McMaster University, Hamilton, Ontario, Canada

Correspondence to
Dr Ariel Izcovich;
ariel.izcovich@gmail.com

ABSTRACT

Introduction Using the best current evidence to inform clinical decisions remains a challenge for clinicians. Given the scarcity of trustworthy clinical practice guidelines providing recommendations to answer clinicians' daily questions, clinical decision support systems (ie, assistance in question identification and answering) emerge as an attractive alternative. The trustworthiness of the recommendations achieved by such systems is unknown.

Objective To evaluate the trustworthiness of a question identification and answering system that delivers timely recommendations.

Design Cross-sectional study.

Methods We compared the responses to 100 clinical questions related to inpatient management provided by two rapid response methods with 'Gold Standard' recommendations. One of the rapid methods was based on PubMed and the other on Epistemonikos database. We defined our 'Gold Standard' as trustworthy published evidence-based recommendations or, when unavailable, recommendations developed locally by a panel of six clinicians following the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach. Recommendations provided by the rapid strategies were classified as potentially misleading or reasonable. We also determined if the potentially misleading recommendations could have been avoided with the appropriate implementation of searching and evidence summary tools.

Results We were able to answer all of the 100 questions with both rapid methods. Of the 200 recommendations obtained, 6.5% (95% CI 3% to 9.9%) were classified as potentially misleading and 93.5% (95% CI 90% to 96.9%) as reasonable. 6 of the 13 potentially misleading recommendations could have been avoided by the appropriate usage of the Epistemonikos matrix tool or by constructing summary of findings tables. No significant differences were observed between the evaluated rapid response methods.

Strengths and limitations of this study

► The study was carried out in a real-world scenario (questions related to patients being treated in a clinical ward).
► Three different clinicians were randomly assigned to apply the different answering strategies.
► We developed a transparent framework to categorise the recommendations obtained by the rapid strategies.
► We sought to provide trustworthy 'Gold Standard' recommendations; nevertheless, it is not possible to guarantee that they were optimal.
► It is unclear if the observed results can be replicated in other settings, for example, with participants less trained in evidence-based decision-making.

Conclusion A question answering service based on the GRADE approach proved feasible to implement and provided appropriate guidance for most identified questions. Our approach could help stakeholders in charge of managing resources and defining policies for patient care to improve evidence-based decision-making in an efficient and feasible manner.

INTRODUCTION

Research consistently shows that there is an important gap between evidence and practice,[1 2] and clinicians seldom use the best available evidence to guide their decisions.[3–5] Limited time, lack of training in critical appraisal and low expectations for finding relevant answers are among the most commonly identified obstacles.[6 7] These practices are problematic as the benefits of using the best current evidence to inform clinical decisions are widely accepted to such

extent that evidence-based decision-making is frequently considered a measure of healthcare quality.[8] In particular, hospital executive boards, insurance companies and consumers recognise that evidence-based practice may help prevent unsafe or inefficient practices.[9–11]

One of the potential solutions for bringing evidence to bedside decisions is the use of trustworthy and transparent clinical practice guidelines. Although the last decade has seen significant advances in guideline methodology (http://www.gradeworkinggroup.org/), important limitations still remain: (1) only a small number of guidelines have been tailored to clinicians' needs;[12] (2) finding relevant guidelines can be laborious and time consuming; and (3) typically, only a few guidelines are kept up to date.[13]

Another alternative for bridging the gap between evidence and clinical practice are clinical decision support systems designed to provide assistance to clinicians in the question identification and resolution process by finding the answer for them and presenting the information in a user-friendly way.[14–18] Unlike products that passively provide preappraised evidence at the point of care (eg, UpToDate), this systems involve trained practitioners that search and deliver tailored answers to identified questions. However, the trustworthiness of the recommendations achieved by such systems is unknown.

The objective of this study was to evaluate the trustworthiness of a question identification and answering system that delivers timely recommendations to clinicians providing care to inpatients by comparing the imparted guidance with 'Gold Standard' recommendations. Additionally, we come up with a proposal on how to replicate the process.

METHODS

We conducted the study on the Internal Medicine Service of the German Hospital of Buenos Aires, Argentina, from March 2014 to March 2016. The context in which this study was carried out has been described in another publication.[17]

We compared two rapid response methods with trustworthy published evidence-based recommendations or, when not available, recommendations developed locally by a panel of six clinicians which, for the purpose of this study, we considered as our 'Gold Standard'. One of the rapid response methods was based on PubMed using clinical queries, which are a series of filters designed to improve the retrieval of scientifically strong and clinically relevant articles from PubMed database.[19] The other was based on Epistemonikos, which is a relational, collaborative, multilingual database of health evidence that includes systematic reviews from multiple sources (Cochrane database of systematic reviews and PubMed, among others).[20]

Three clinicians trained in evidence-based decision-making (informationists) attempted to answer all the identified questions following three different strategies.

The informationists differ from clinical librarians in that they are trained in clinical epidemiology methods rather than simply information acquisition and have clinical expertise relevant to the questions that allows contextual interpretation of research findings. Each question had its own randomisation schedule drawn from a computer pregenerated random number list in which each informationist was assigned to one of the three strategies defined below. We describe the question identification process and the strategies to address the questions in the following sections.

Identification and selection of clinical questions

One of the informationists (AI), otherwise uninvolved in the patients' care, identified questions relevant to the staff and residents of the Internal Medicine Service. Either the staff or residents explicitly formulated the questions, or AI inferred them from the discussion of the clinical cases. He collected the relevant clinical question using the Population/Problem, Intervention, Comparison, Outcome (PICO) framework.

In order to focus on questions that could potentially impact clinicians' course of action, we excluded questions that (1) were answered immediately by someone who was present in the session, other than the informationists, typically using electronic resources such as UpToDate; (2) were not related to therapeutic or diagnostic interventions and (3) addressed interventions already implemented in the patient's care.

All the identified questions that did not fulfilled one of the exclusion criteria were included and registered. The described question identification process was repeated until the study was finished.

Rapid strategy based on PubMed (strategy 1)

The informationist assigned to this strategy performed a literature search on MEDLINE using the PubMed clinical queries feature (online supplementary figure 1). First, he tried to identify relevant systematic reviews;[21] when unavailable or when considered that additional relevant information could exist, he also searched for primary studies. Once the informationist identified the most relevant systematic review and/or primary study/s, he followed the GRADE approach to interpret the results and judge the certainty on the evidence (for a detailed description, see Grading of Recommendations Assessment, Development and Evaluation (GRADE) handbook available at http://gdt.guidelinedevelopment.org/app/handbook/handbook.html). Following the GRADE guidance, the informationist also considered additional relevant information related to patients' values and preferences, costs, applicability and feasibility,[22 23] and made a clinical decision simulating what clinicians could do in the optimal scenario. To capture the decision, the informationist formulated a recommendation that included the direction (in favour or against the intervention) and the strength (strong or weak). The process took no more than 2 hours.

Rapid strategy based on Epistemonikos (strategy 2)

The informationist assigned to this strategy searched on the Epistemonikos database using the 'matrices of evidence' tool, which is a tabular way of displaying the cluster of systematic reviews that share at least one included study[24] and followed the same process described for strategy 1 (online supplementary figure 2). He also searched PubMed for randomised controlled trials (RCT) in cases where systematic reviews were not available or when he considered that additional relevant information could exist (online supplementary figure 2).

Strategy based on trustworthy recommendations ('Gold Standard') (strategy 3)

The informationist assigned to this strategy searched for recommendations developed with the GRADE approach on the following databases: Tripdatabase (http://www.tripdatabase.com); National guideline Cleringhouse (http://www.guidelines.gov); Canadian Medical Association (https://www.cma.ca); National Institute for Health and Care Excellence (http://www.nice.org.uk/); SIGN (http://www.sign.ac.uk); GuíaSalud (http://portal.guiasalud.es/web/guest/buscar-gpc); Australian clinical practice guidelines (http://www.clinicalguidelines.gov.au); New Zealand Guidelines Group (http://www.nzgg.org.nz/); US preventive Task Force (http://www.uspreventiveservicestaskforce.org/); eGuidelines (https://www.guidelines.co.uk/) and GIN (http://www.g-i-n.net/about-g-i-n/introduction).

He critically assessed the identified recommendations using the criteria proposed for evaluating GRADE recommendations[25] and qualitatively categorised their trustworthiness as high, moderate or low based on the answers to the following questions: Was the question clearly formulated? Were all the critical outcomes considered? Was the recommendation based on the best current evidence? The evidence was clearly presented? Was the recommendation coherent with the supporting evidence? Were the values and preferences considered?

Additionally, for every question, the same informationist searched for systematic reviews, RCTs and observational studies on the following databases without time restriction: PubMed, Epistemonikos and the Cochrane database of systematic review. He used the information extracted from the relevant systematic reviews and/or primary studies to construct a Summary of Finding Table (SoF) following the GRADE principles (SoF example available in online supplementary table 1).[26 27] The tables were then sent via email to six clinicians ('local panel') with experience in applying the GRADE approach. Each clinician used the information included in the SoF tables and considered issues related to patients' values and preferences, costs, applicability and feasibility to individually construct a recommendation.[22 23] When >66% of the clinicians who answered agreed on the strength and direction of the recommendation, we considered that recommendation final. Disagreement in the direction or the strength of the recommendation was recorded and resolved by seventh clinician (IN) with experience in developing GRADE recommendations. Although we intended to answer every question with the described 'local panel' approach, we only used the resultant recommendations when published GRADE recommendations developed by guideline panels rated as 'high' for trustworthiness
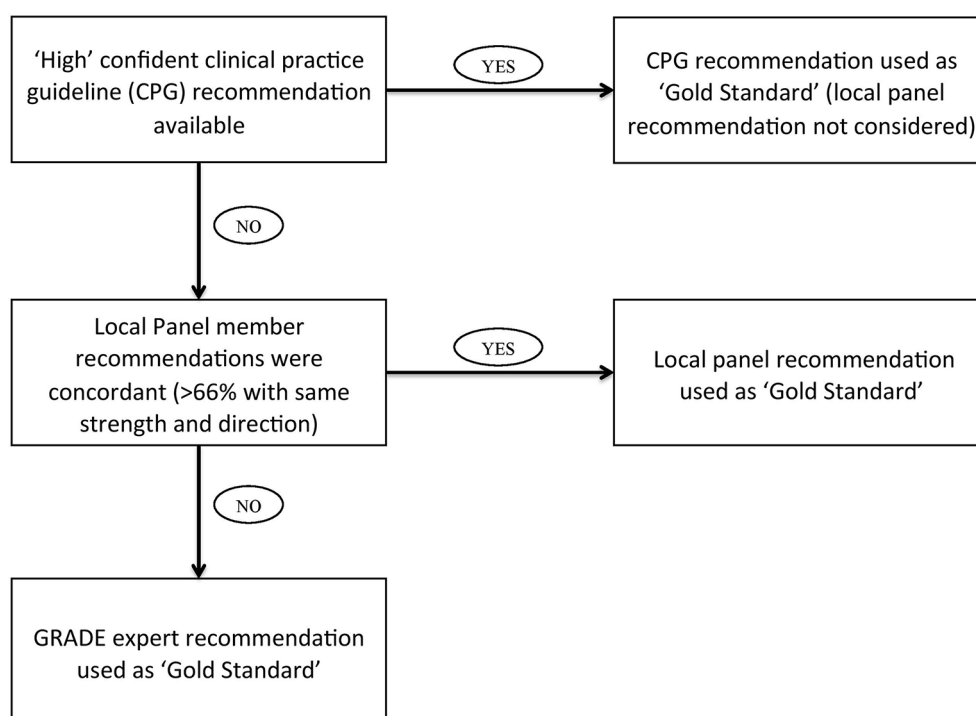

Figure 1 'Gold Standard' recommendation development.

were unavailable. Figure 1 provides a description of the 'Gold Standard' recommendation construction process.

Outcomes

We compared the recommendations, quality of evidence judgements and information used by rapid strategies and the 'Gold Standard' strategy to define the following outcomes:

- ► Inappropriate recommendations: when the 'Gold Standard' was a strong recommendation and the rapid strategies yielded a decision in the opposite direction of any strength or when the 'Gold Standard' was a weak recommendation and the rapid strategies yielded a strong recommendation in the opposite direction.
- ► Overconfident recommendations: when the 'Gold Standard' was a weak recommendation and the rapid strategies yielded a decision concordant with a strong recommendation on the same direction.
- ► Potentially misleading recommendations: composite of inappropriate or overconfident recommendations.
- ► Concordant recommendations: when the 'Gold Standard' and the rapid strategies yielded a recommendation of the same direction and strength
- ► Reasonable disagreement: when the 'Gold Standard' was a weak recommendation in favour and the rapid strategies yielded a weak recommendation against or vice versa, or when the 'Gold Standard' was a strong recommendation and the rapid strategies yielded a weak recommendation on the same direction
- ► Reasonable recommendations: composite of concordant recommendations and reasonable disagreement.
- ► Same direction recommendations: when the Gold Standardandthe rapid strategies yielded a recommendation of the same direction regardless of its strength.
- ► Inappropriate quality of evidence judgement: proportion of recommendations in which the quality of evidence (1) was judged as low or very low by the rapid strategies and high or moderate by the 'Gold Standard' or (2) was judged as high or moderate by the rapid strategies and low or very low by the 'Gold Standard'.
- ► Coincidence in information usage: proportion of recommendations in which the publications used by the rapid methods were the same as the ones used the 'Gold Standard'.

Table 1 describes the framework for rapid recommendation categorisation based on their comparison with 'Gold Standard' recommendations.

Additional analyses

We also performed a post hoc qualitative analysis of the recommendations classified as potentially misleading. We analysed the reasons for the disagreement between the rapid strategies and the gold standard and we considered potential solutions. For this purpose, in cases in which the potentially misleading recommendations were judged to be a consequence of inadequate evidence selection, we determined if the appropriate use of the Epistemonikos matrices tool could have prevented that problem (ie, identification of a systematic review containing primary studies that were not considered for the development of the original recommendation). In cases in which potentially misleading recommendations were judged to be a consequence of inappropriate evidence interpretation, we determined if the correct presentation of the evidence could have prevented the problem. To assess this, we sent the SoF table constructed in response to the same question for the 'Gold Standard' strategy (strategy 3) to the investigator who originally constructed the potentially misleading recommendation. We asked the investigator to provide a new recommendation based in the SoF. We judged that the correct use of the SoF could have prevented the problem when the investigator provided a reasonable recommendation in response (compared with the Gold Standard (GS) recommendation).

Statistical analysis

For the comparisons between the rapid strategies and the 'Gold Standard', we calculated proportions and 95% CI for all the outcomes. We also calculated inter-rater agreement with kappa statistic using VassarStats calculator (http://vassarstats.net/kappa.html). For the kappa calculation related to recommendation concordance (strong in favour, weak in favour, weak against or strong against), we imputed the double of distance between strong in favour—weak in favour and strong against—weak against than weak in favour—weak against. For the kappa calculation related to quality of evidence agreement (high, moderate, low or very low), we imputed the double of distance between moderate—low than very low—low and moderate—high. For the comparison between strategies

Table 1	Framework to categorise recommendations			
	Gold Standard			
	Strong against	**Weak against**	**Weak in favour**	**Strong in favour**
Rapid strategies				
Strong against	Concordant	Overconfident	Inappropriate	Inappropriate
Weak against	Reasonable	Concordant	Reasonable	Inappropriate
Weak in favour	Inappropriate	Reasonable	Concordant	Reasonable
Strong in favour	Inappropriate	Inappropriate	Overconfident	Concordant

Table 2 Recommendations according to the strategy implemented

	Strategy 1 (n=100)	Strategy 2 (n=100)	Strategy 3 (CPG) (n=80) (%)	Strategy 3 ('local panel') (n=100)
Recommendations				
Strong	14	12	21 (26.2)	21
In favour of the intervention	55	62	55 (68.7)	63
Quality of evidence				
High	8	5	–	12
Moderate	22	25	–	28
Low	34	26	–	44
Very low	36	44	–	16
Confidence in the CPG recommendation				
High (%)	–	–	16 (20)	–

CPG, clinical practice guideline.

RESULTS

1 and 2, we calculated relative risks and 95% CI when possible.

During the study period, we identified 100 questions all of which were answered with strategies 1 and 2 (200 recommendations). With strategy 3, we found recommendations in clinical practice guideline (CPG) for 80 of the 100 questions all of which could be answered by the 'local panel' approach. The process of answering each question with strategy 3 ('Gold Standard, local panel' approach) took, on average, 1 week per question. Table 2 presents the characteristics of the recommendations delivered by each strategy. A list of the PICOs is available in the online supplementary table 2.

Following the process described in figure 1, we obtained 100 'Gold Standard' recommendations. These recommendations were composed by 16 high confidence CPG recommendations, 55 panel recommendations and 29 expert recommendations. The results of the comparison between the rapid strategies and the 'Gold Standard' are described in table 3.

The comparison between strategies 1 and 2 is described in online supplementary table 3.

There were 13 recommendations that were judged as potentially misleading; the causes and possible solutions are summarised in online supplementary table 4 .

Table 3 Rapid strategies recommendations analysis

	Rapid strategies versus 'Gold Standard' (n=200)	Kappa
Potentially misleading recommendations	6.5% (3%–9.9%)	–
Inappropriate	3.5% (0.95%–6%)	–
Overconfident	3% (0.64%–5.3%)	–
Reasonable recommendations	93.5% (90%–96.9%)	–
Concordant	62.5% (55.7%–69.2%)	0.59 (0.36–0.82)
Reasonable disagreement	31% (24.5%–37.4%)	–
Same direction recommendations	74% (67.5%–79.5%)	–
Strong (rapid strategies) (n=26)	96.1% (82.2%–99.3%)	–
Weak (rapid strategies) (n=174)	70.6% (64.5%–76.9%)	–
Potentially misleading quality of evidence judgement	20% (14.4%–25.5%)	–
Inappropriate moderate or high	5% (1.9%–8%)	–
Inappropriate low or very low	15% (10%–19.9%)	–
Quality of evidence agreement	55.5% (48.6%–62.3%)	0.59 (0.46–0.72)
Coincidence in information use*	60% (50.4–69.6)	–

*The same publication/s were used to answer the question.

DISCUSSION

The results of the present study suggest that a rapid question answering system based on the GRADE approach provided appropriate guidance in response to most questions. Although the proportion of concordant recommendations (same strength and direction between rapid strategies and GS) was 62.5%, most of the remainder (31% of the total) were classified as 'reasonable disagreements'. Only 13 of the 200 recommendations were judged as potentially misleading and approximately half of those could possibly have been avoided with an appropriate use of the available tools (Epistemonikos matrix of evidence or SoF tables). We also analysed the results considering exclusively the direction of the recommendations. The results showed that almost all strong recommendations constructed with the rapid strategies shared the same 'Gold Standard's' direction while 70% of the weak recommendations did. This finding is not surprising given that weak recommendations are frequently based on low or very low quality of evidence, or are warranted in situations where benefits and risks are closely balanced, hence their direction is subjectively defined by weighting those aspects (eg, in a setting in which benefits and harms are balanced, some guideline panel members can interpret the results as favouring the intervention while others as favouring the comparison).[22 23 25] Although 30% of weak recommendations had a different direction from the 'Gold Standard's', we consider that it is unlikely that they would have resulted in misleading guidance as those willing to use them should carefully analyse the fundamentals of the recommendation before deciding their course of action.[22–24] An exception would be the situation in which the 'Gold Standard' recommendations were strong in the opposite direction, but this was captured in the primary analysis as those recommendations were classified as inappropriate. A third analysis in which we calculated rapid strategies' and 'Gold Standard's' recommendation strength and direction agreement beyond chance, using weighted kappa, informed moderate to substantial agreement.[28] As described for the former analysis (considering only the direction of recommendations), this approach also does not acknowledge the possibility of reasonable disagreement. Hence, it only reflects the capability of the rapid strategies to provide concordant recommendations (same direction and strength) with the 'Gold Standard's' which we believe is an overdemanding approach that underestimates the ability of the rapid strategies to provide adequate guidance.

The comparison between the different rapid answering strategies (PubMed vs Epistemonikos) showed that although the proportion of potentially misleading recommendations was small in both strategies, there was a slight (3%) absolute difference in favour of PubMed strategy. One possible explanation for the difference is that the investigators involved in the study were less familiarised with Epistemonikos database and search engine than PubMed's.

The main limitation of our study is that it is not possible to define a 'Gold Standard' recommendation for a medical question. We sought to provide trustworthy 'Gold Standard' recommendations by performing rigorous evidence searches, constructing detailed evidence summaries and including multiple clinicians trained in evidence-based decision-making; nevertheless, this approach does not guarantee optimal recommendations. In addition, the system was applied to a specific subgroup of questions (intervention-related questions that were not immediately answered). We consider that addressing questions that do not meet these criteria are less likely to change clinicians' behaviour. Also this study was carried out in a singular context (clinicians trained in evidence-based decision-making with advanced understanding of the GRADE system). It is unknown to what extent the observed results can be replicated in different situations where clinicians are less familiarised with evidence-based medicine concepts.

Although investigators have previously undertaken evaluation of the implementation of question answering services,[29–34] these studies focused on clinicians' attitudes and decisions in response to the answers provided. As long as it remains uncertain that the answers the services provide are based on the best available evidence, and that clinicians interpret and use the provided information appropriately to make coherent decisions, the benefits of the implementation of these services to improve patient outcomes cannot be assumed.[35] Another approach would be to directly measure the impact of answering clinicians' questions on patients' clinical important outcomes (ie, mortality or length of hospital stay). However, for these kinds of interventions that are designed to improve quality of care through affecting physician's behaviour, demonstrating such an effect could be very difficult (huge sample sizes needed, low signal-to-noise ratio).[36–38] Attempts have been made in this direction and the results suggest possible benefits with the implementation of the evaluated interventions but the quality of evidence provided was low, either because of imprecision (underpowered studies)[16–18] or because of risk of bias (non-randomised comparisons).[39–41]

We found only one study that considered the trustworthiness of the answers provided.[42] In that study, the investigators inserted study evidence statements related to the management of clinical conditions for which high-quality RCTs or meta-analyses had unequivocally established benefits greater than risks, costs and inconvenience into hospital discharge letters. The study results showed a significant increase in general practitioner adherence to discharge medications demonstrating that, in optimal conditions (no time restrictions to perform evidence searches, high quality of evidence available), providing information to clinicians improve patient care. However, that optimal scenario is probably the exception as for most clinical questions high-quality evidence remains unavailable[16 43 44] and clinicians usually need very prompt answers to their questions. Hence, ours is

the first study to use a structured and objective approach to measure the quality of the information provided in a timely way to clinician-generated questions.

To achieve a medical practice consistent with what Ubbink *et al* described as evidence-based practice,[45] clinicians need to be able to quickly obtain and accurately assess the best available evidence to answer their questions. Clinical practice guidelines endeavour to provide these answers at the point of care and, when rigorously developed and up to date, constitute optimal guidance. However, most of the available guidelines have methodological flaws and do not provide trustworthy recommendations.[12 13] In the present study, 80% of the identified questions could be answered with recommendations included in CPG but only 20% of them were judged to be trustworthy.

Given current guideline limitations, if feasible and properly implemented, a question answering system could provide a solution. This study adds to our previous study in which we evaluated the impact of implementing a response system, similar to the one evaluated in the preset trial, on clinicians' decisions.[16] The results of that trial suggested that these kinds of interventions can influence clinicians' courses of actions and therefore patient care.

The present study was developed in a real-life scenario with limited amount of resources, which suggest that the proposed intervention can possibly be implemented in a variety of settings, including a busy clinical ward. We were able to efficiently implement the proposed system with (1) one clinician trained in evidence-based decision-making exclusively dedicated to this task for at least 2 hours a day and (2) a computer with internet connection. We used a systematic and transparent method to arrive at decisions. Finally, we have developed a framework to compare different recommendations developed with the GRADE approach acknowledging that not every discrepancy should be considered inappropriate as different values and preferences may lead to reasonable disagreement between recommendations.

Implication for practice

Those interested in improving evidence use in healthcare decision-making should consider the implementation of systems as the one proposed in the present study. This would require, at least, one trained healthcare provider (informationist) who would (1) search for trustworthy published recommendations or, when not available, systematic reviews in Epistemonikos and/or PubMed; (2) use the Epistemonikos matrices of evidence tool and/or PubMed to identify additional information (not included in the selected systematic review); (3) construct a summary of findings table including all critical outcomes and (4) define a recommendation based on the identified trustworthy recommendations or the summary of findings tables (figure 2). We think that the cornerstone
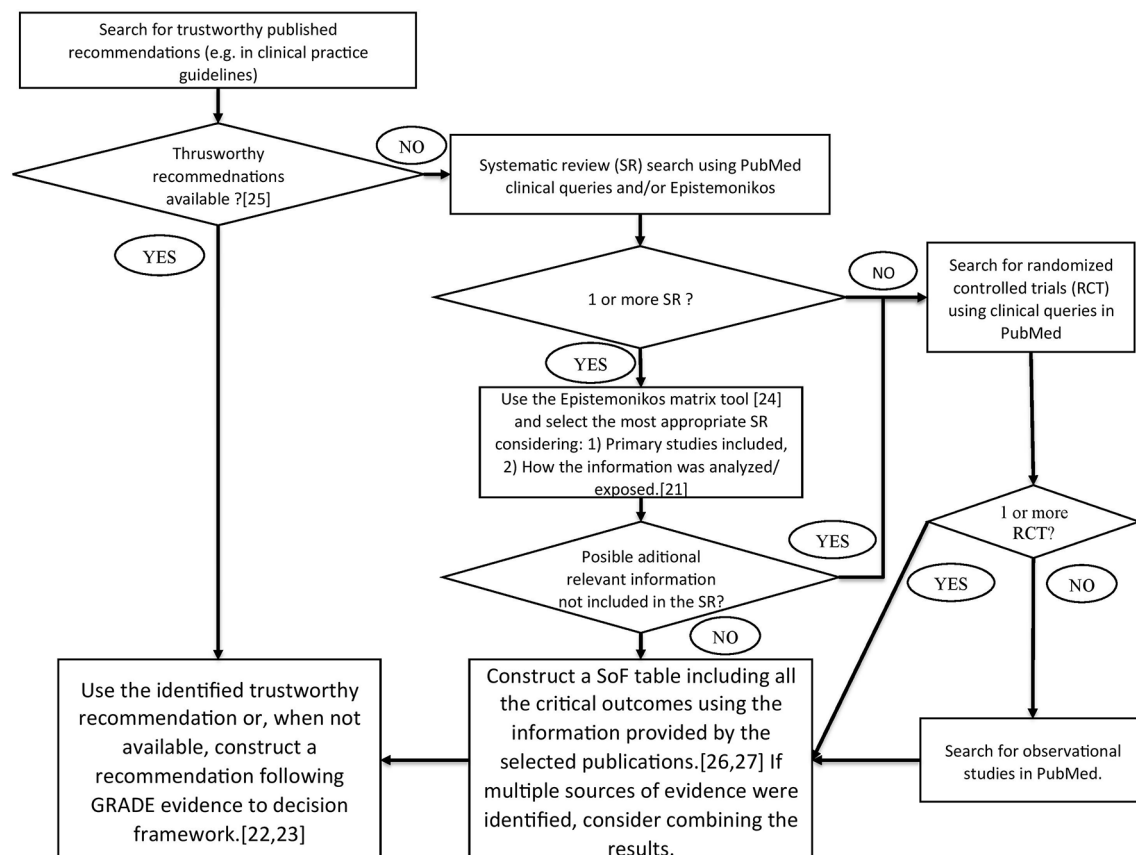

Figure 2 Rapid answering system proposal. SoF, Summary of Finding Table.

to successfully replicate the described process is practitioners training in evidence search, critical appraisal, summary and evidence to decision translation.

Implication for research

Investigators who addressed the clinical questions using the proposed strategies in the present study were highly trained in evidence-based decision-making and could possibly be classified as experts. Whether similar results could be obtained when those responsible for solving the identified questions are not experts remains uncertain.

CONCLUSION

A question answering service based on the GRADE approach proved feasible to implement and provided appropriate guidance for most identified questions. Our approach could help stakeholders in charge of managing resources and defining policies for patient care to improve evidence-based decision-making in an efficient and feasible manner.

REFERENCES

1. Committee on Quality of Health Care in America. *Institute of Medicine. crossing the quality chasm: a new health system for the 21st century*. Washington DC: National Academies Press, 2001.
2. Mickan S, Burls A, Glasziou P. Patterns of 'leakage' in the utilisation of clinical guidelines: a systematic review. *Postgrad Med J* 2011;87:670–9.
3. Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. *Health Info Libr J* 2007;24:78–94.
4. Dawes M, Sampson U. Knowledge management in clinical practice: a systematic review of information seeking behavior in physicians. *Int J Med Inform* 2003;71:9–15.
5. Coumou HC, Meijman FJ. How do primary care physicians seek answers to clinical questions? A literature review. *J Med Libr Assoc* 2006;94:55–60.
6. Scott I, Heyworth R, Fairweather P, *et al*. The use of evidence-based medicine in the practice of consultant physicians. Results of a questionnaire survey. *Aust N Z J Med* 2000;30:319–26.
7. Ely JW, Osheroff JA, Chambliss ML, *et al*. Answering physicians' clinical questions: obstacles and potential solutions. *J Am Med Inform Assoc* 2005;12:217–24.
8. Alexander JA, Hearld LR. What can we learn from quality improvement research? A critical review of research methods. *Med Care Res Rev* 2009;66:235–71.
9. Glasziou P, Ogrinc G, Goodman S. Can evidence-based medicine and clinical quality improvement learn from each other? *BMJ Qual Saf* 2011;20 Suppl 1:i13–i17.
10. Otero HJ, Chambers JD, Bresnahan BW, *et al*. Medicare's national coverage determinations in diagnostic radiology: examining evidence and setting limits. *Acad Radiol* 2012;19:1060–5.
11. Balakas K, Potter P, Pratt E, *et al*. Evidence Equals Excellence: The application of an evidence-based practice model in an academic medical center. *Nurs Clin North Am* 2009;44:1–10.
12. Alonso-Coello P, Irfan A, Solà I, *et al*. The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. *Qual Saf Health Care* 2010;19:e58.
13. Martínez García L, Sanabria AJ, García Alvarez E, *et al*. The validity of recommendations from clinical guidelines: a survival analysis. *CMAJ* 2014;186:1211–9.
14. Swinglehurst DA, Pierce M, Fuller JC. A clinical informaticist to support primary care decision making. *Qual Health Care* 2001;10:245–9.
15. Verhoeven AA, Schuling J. Effect of an evidence-based answering service on GPs and their patients: a pilot study. *Health Info Libr J* 2004;21(Suppl 2):27–35.
16. Izcovich A, Criniti JM, Ruiz JI, *et al*. Impact of a GRADE-based medical question answering system on physician behaviour: a randomised controlled trial. *Evid Based Med* 2015;20:81–7.
17. Izcovich A, Malla CG, Diaz MM, *et al*. Impact of facilitating physician access to relevant medical literature on outcomes of hospitalised internal medicine patients: a randomised controlled trial. *Evid Based Med* 2011;16:131–5.
18. Pastori MM, Sarti M, Pons M, *et al*. Assessing the impact of bibliographical support on the quality of medical care in patients admitted to an internal medicine service: a prospective clinical, open, randomised two-arm parallel study. *Evid Based Med* 2014;19:163–8.
19. Haynes RB, Wilczynski N. Finding the gold in MEDLINE: clinical queries. *ACP J Club* 2005;142:101–2.
20. Rada G, Pérez D, Capurro D, *et al*. Epistemonikos: a free, relational, collaborative, multilingual database of health evidence. *Stud Health Technol Inform* 2013;192:486–90.
21. Murad MH, Montori VM, Ioannidis JP, *et al*. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. *JAMA* 2014;312:171–9.
22. Andrews J, Guyatt G, Oxman AD, *et al*. GRADE guidelines: 14. going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol* 2013;66:719–25.
23. Andrews JC, Schünemann HJ, Oxman AD, *et al*. GRADE guidelines: 15. going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol* 2013;66:726–35.
24. Rada G, Neumann I, Cerda J. A new way to find evidence quickly and efficiently. *Medwave* 2014;14:e6044.
25. Neumann I, Santesso N, Akl EA, *et al*. A guide for health professionals to interpret and use recommendations in guidelines developed with the GRADE approach. *J Clin Epidemiol* 2016;72:45–55.
26. Guyatt GH, Oxman AD, Santesso N, *et al*. GRADE guidelines: 12. preparing summary of findings tables-binary outcomes. *J Clin Epidemiol* 2013;66:158–72.
27. Guyatt GH, Thorlund K, Oxman AD, *et al*. GRADE guidelines: 13. preparing summary of findings tables and evidence profiles-continuous outcomes. *J Clin Epidemiol* 2013;66:173–83.
28. McGinn T, Wyer PC, Newman TB, *et al*. Tips for learners of evidence-based medicine: 3. measures of observer variability (kappa statistic). *CMAJ* 2004;171:1369–73.
29. Brettle A, Maden-Jenkins M, Anderson L, *et al*. Evaluating clinical librarian services: a systematic review. *Health Info Libr J* 2011;28:3–22.
30. Bryant SL, Gray A. Demonstrating the positive impact of information support on patient care in primary care: a rapid literature review. *Health Info Libr J* 2006;23:118–25.
31. Brassey J, Elwyn G, Price C, *et al*. Just in time information for clinicians: a questionnaire evaluation of the ATTRACT project. *BMJ* 2001;322:529–30.
32. McGowan J, Hogg W, Campbell C, *et al*. Just-in-time information improved decision-making in primary care: a randomized controlled trial. *PLoS One* 2008;3:e3785.
33. Marshall JG, Neufeld VR. A randomized trial of librarian educational participation in clinical settings. *J Med Educ* 1981;56:409–16.
34. Mulvaney SA, Bickman L, Giuse NB, *et al*. A randomized effectiveness trial of a clinical informatics consult service: impact on evidence-based decision-making and knowledge implementation. *J Am Med Inform Assoc* 2008;15:203–11.

35. Guyatt GH, Meade MO, Jaeschke RZ, *et al*. Practitioners of evidence based care. Not all clinicians need to appraise evidence from scratch but all need some skills. *BMJ* 2000;320:954–5.

36. Batalden PB, Davidoff F. What is "quality improvement" and how can it transform healthcare? *Qual Saf Health Care* 2007;16:2–3.

37. Glasziou P. What is EBM? *Evid Based Med* 2011;16:129–30.

38. Mant J, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ* 1995;311:793–6.

39. Isaac T, Zheng J, Jha A. Use of UpToDate and outcomes in US hospitals. *J Hosp Med* 2012;7:85–90.

40. Bonis PA, Pickens GT, Rind DM, *et al*. Association of a clinical knowledge support system with improved patient safety, reduced complications and shorter length of stay among Medicare beneficiaries in acute care hospitals in the United States. *Int J Med Inform* 2008;77:745–53.

41. Banks DE, Shi R, Timm DF, *et al*. Decreased hospital length of stay associated with presentation of cases at morning report with librarian support. *J Med Libr Assoc* 2007;95:381–7.

42. Kunz R, Wegscheider K, Guyatt G, *et al*. Impact of short evidence summaries in discharge letters on adherence of practitioners to discharge medication. A cluster-randomised controlled trial. *Qual Saf Health Care* 2007;16:456–61.

43. Alexander PE, Bero L, Montori VM, *et al*. World Health Organization recommendations are often strong based on low confidence in effect estimates. *J Clin Epidemiol* 2014;67:629–34.

44. Djulbegovic B, Kumar A, Kaufman RM, *et al*. Quality of evidence is a key determinant for making a strong GRADE guidelines recommendation. *J Clin Epidemiol* 2015;68:727–32.

45. Ubbink DT, Guyatt GH, Vermeulen H. Framework of policy recommendations for implementation of evidence-based practice: a systematic scoping review. *BMJ Open* 2013;3:e001881.