



# Impact of a GRADE-based medical question answering system on physician behaviour: a randomised controlled trial

Ariel Izcovich, Juan Martín Criniti, Juan Ignacio Ruiz, Hugo Norberto Catalano

10.1136/ebmed-2014-110146

► Additional material is published online. To view please visit the journal (<http://dx.doi.org/10.1136/ebmed-2014-110146>)

Internal Medicine Department, Hospital Alemán de Buenos Aires, Buenos Aires, Argentina

Correspondence to: **Dr Ariel Izcovich**, Internal Medicine Department, Hospital Alemán de Buenos Aires, Pueyrredon 1640, Buenos Aires 1118, Argentina; [ariel.izcovich@gmail.com](mailto:ariel.izcovich@gmail.com)



Editor's choice  
Scan to access more  
free content

## Abstract

Physicians are frequently faced with questions related to their patients' care that they cannot answer. A vast number of randomised trials have tested a wide variety of behaviour-changing strategies designed to improve practitioners' evidence utilisation, but systematic reviews have concluded that the effects are generally small and inconsistent. We conducted a randomised controlled trial to determine whether a question identification and solving system, using structured evidence summaries with recommendations, would change physician's behaviour related to the care of their hospitalised patients. The trial was conducted at the secondary level, internal medicine ward. Relevant clinical questions were the units of randomisation; 14 clinicians participated in the study. The question identification and answering system was carried out using evidence summaries with recommendations based on the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) approach stressing influence on clinician behaviour (decision/recommendation concordance). During 131 morning reports, 553 questions were identified (4.2 questions per meeting). 398 were excluded because they were not about diagnostic or therapeutic interventions or because their answers could not have impact on clinician behaviour, and 31 were excluded because of lack of time to answer them, leaving 124 included questions. The proportion of clinical decisions concordant with the proposed recommendations was 79% in the intervention arm and 44% in the control arm: relative risk 1.8 (95% CI 1.3 to 2.4), number of evidence summaries needed to change a care decision for one question raised was 3 (95% CI 2 to 6). A question identification and answering system was feasible, effectively performed and significantly influenced clinician behaviour related to the care of hospitalised patients, which suggests that interventions facilitating accessibility and interpretability of the best available evidence at the point of care have the potential to significantly impact on the quality of healthcare.

## Introduction

Physicians are frequently confronted with questions related to their patients' care<sup>1</sup> that they cannot answer.<sup>2,3</sup> Several reasons for this behaviour have been identified and include limited time to perform the research, lack of training in critical appraisal of the information found and low expectations for finding a relevant and direct answer to questions.<sup>4,5</sup> Improving this situation would mean a step towards achieving clinical practice that is consistent with the best evidence. A vast number of randomised trials have tested a wide variety of behaviour-change strategies designed to improve practitioners' evidence utilisation, but

systematic reviews have concluded that the effects are generally small and inconsistent.<sup>6,7</sup> Initiatives aimed to allow physicians to have access to up-to-date, evidence-based treatment and management recommendations at the point of care are being developed.<sup>8</sup> Another intuitively appealing way to achieve such evidence-based practice (EBP) is to provide a service of question identification and answering. Such services have shown to be able to provide satisfactory answers,<sup>9</sup> change physician's information-seeking behaviour<sup>10</sup> and increase general practitioner adherence to prescribing discharge medications.<sup>11</sup> We conducted a randomised controlled trial (RCT) to determine if a question identification and solving system using structured evidence summaries with recommendations would influence physician behaviour related to the care of their hospitalised patients.

## Methods

### Medical problem-solving system

The study was conducted on the Internal Medicine Service of a German hospital in Buenos Aires, Argentina, between June 2012 and July 2013, and was approved by the institutional ethics committee (April 2012, ref nr 115). The context in which this study was carried out was described in a previous publication.<sup>12</sup> During the course of the study, a physician specialised in internal medicine, who was trained and skilled in evidence-based decision-making, and whose work was funded by the Internal Medicine Service, identified the medical questions that arose during morning reports in which all the clinicians (staff and residents) in the Internal Medicine Service meet up to discuss cases. Such questions were either explicitly formulated by staff or resident physicians, or inferred. They were collected using the PICOT structure (Population/Problem, Intervention, Comparison, Outcome, Type of design that would answer the question)<sup>13</sup> to gather key words for a literature search. In order to focus on those questions whose answers could potentially change the clinician's course of action, we excluded those that fulfilled one of the following criteria: (1) answered immediately by someone who was present in the session, frequently, using electronic resources such as UpToDate, (2) not related to therapeutic or diagnostic intervention, (3) about past decisions (eg, in the case of a patient with stroke for whom an MRI was performed: would it not have been better to perform a CT scan?). The same physician who collected the questions also subjectively categorised them based on the probability of significant clinical impact on patient important outcomes, and attempted to answer all of them, one by one, prioritising those he considered more relevant. The literature search was carried out following an 'umbrella' strategy in which the first step was the identification of clinical

practice guidelines and systematic reviews,<sup>14</sup> and the second step was the identification of individual RCTs or observations when considered relevant (ie, RCT published after the publication of the identified systematic reviews; figure 1). The search started once the morning report was over and lasted a maximum of 2 h. Using the identified publications, the same clinician analysed all the critical and important outcomes (relative and absolute effects and their confidence levels), judged the benefits/risks balance and its confidence level and translated this information into a recommendation following the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) approach<sup>15 16</sup> (see online supplementary appendix 1). Finally, the clinician constructed short, structured evidence summaries that contained a brief description of the case that motivated the question, a summary of the evidence gathered, an interpretation of the results (description of best estimates of effect and confidence)<sup>17</sup> and the recommendation.

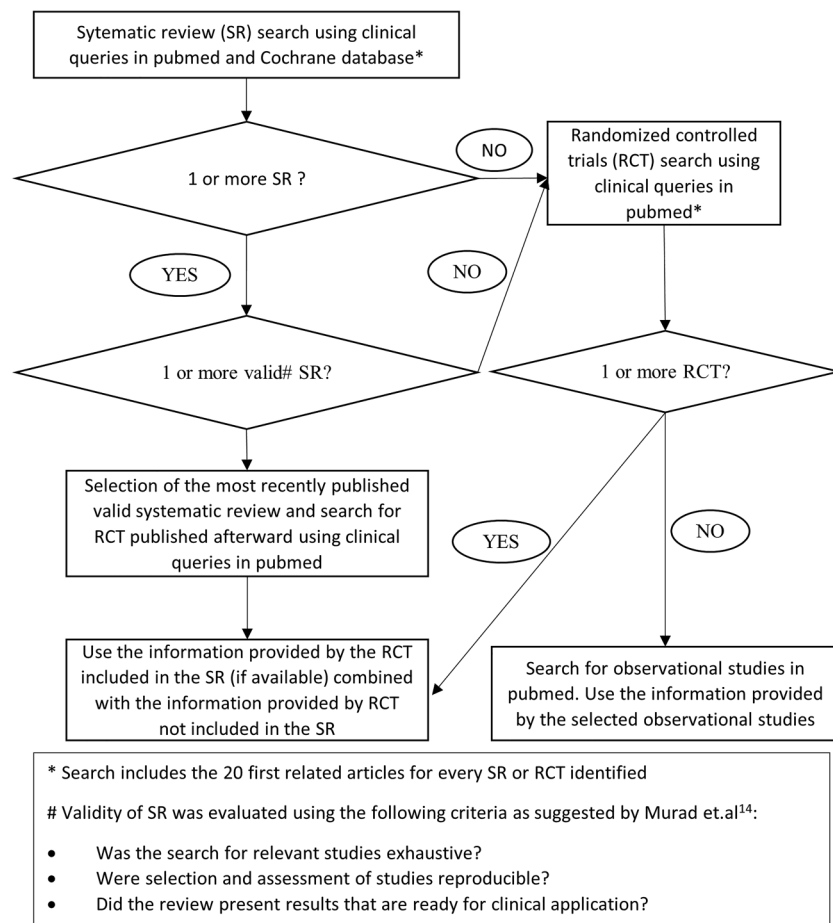
### Study participants

All the Internal Medicine Service attending physicians (six staff and eight residents) participated and were blinded to the purpose and methods of the study until it was over. Although the authors were members of the service during the realisation of the study, they did not attend or made decisions related to patients' care, and were hence not considered as study participants. The

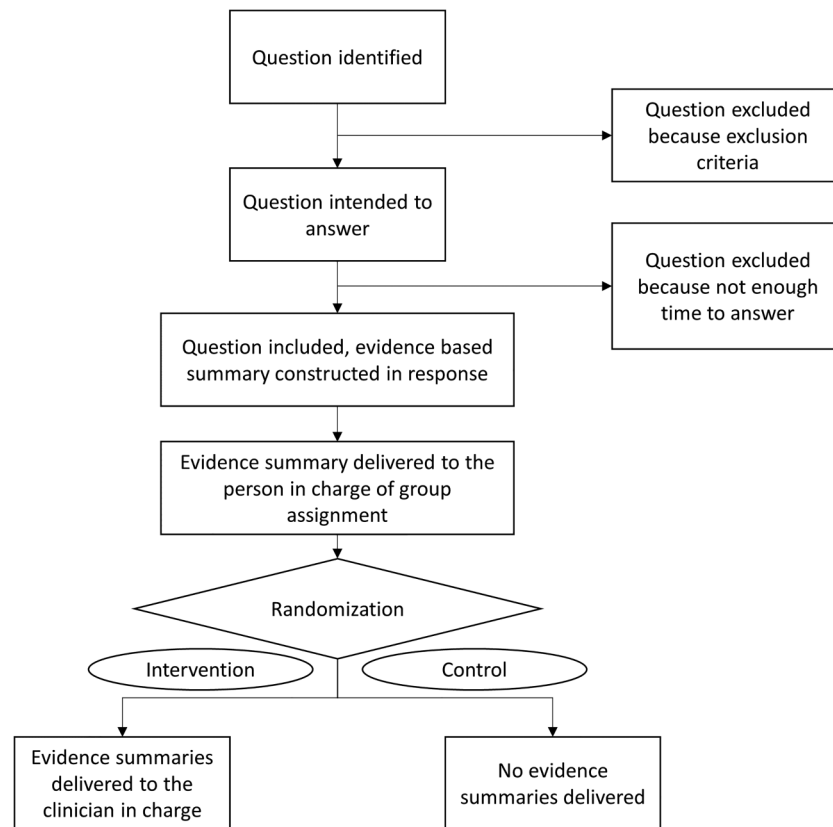
process of the medical question identification and answering described is the usual practice on the service since 2009, which made this possible. All the participating physicians had some degree of training in evidence-based problem solving.

### Study design

After the responsible physician answered the questions as described, they were randomised to two arms, intervention and control. Randomisation was performed using a computer pregenerated random number list and was stratified by strength of recommendation. A member of the service that usually performs administrative tasks, and who was blinded to the study's objectives, was in charge of group assignment. She received all the evidence summaries and delivered (hardcopy and email), to the clinicians in charge, those constructed in response to the questions assigned to the intervention arm. She archived the summaries written in response to questions assigned to the control arm (figure 2). Two clinicians, blinded to arm assignment, checked patient electronic medical records and, when necessary, performed telephonic interviews with patients in order to assess the following outcomes: (1) decision/recommendation concordance (primary outcome) defined as the proportion of medical decisions that were coherent with the proposed recommendations, (2) in hospital mortality, (3) transfer to an intensive care unit (ICU), (4) 6-month mortality,



**Figure 1** Information searching strategy.



**Figure 2** Trial design.

(5) 6-month rehospitalisation. We performed a prespecified subgroup analysis of strong versus weak recommendations (see online supplementary appendix 1).

#### Sample size, power calculation and statistical analysis

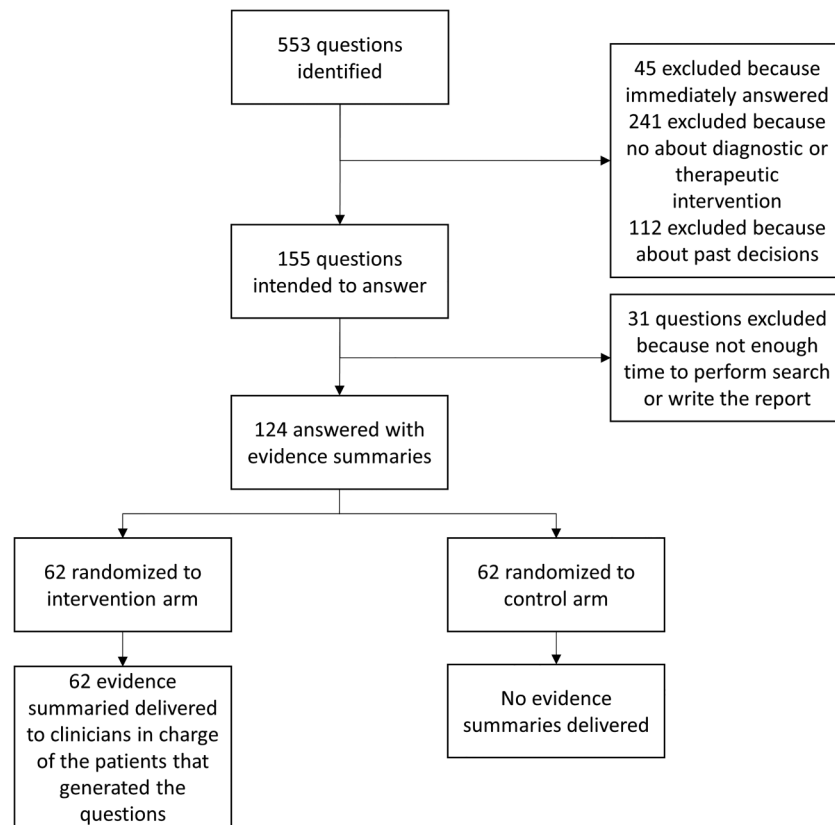
Assuming that without the intervention, half of the clinicians' decisions would be in line with the proposed recommendations, and considering as clinically significant an absolute increase in the compliance of 25% and an  $\alpha$  error of 0.05, a sample size of 116 questions in total would guarantee 80% power. Calculations were performed using G\*power 3 software (<http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3/>).

Dichotomous variables were assessed using relative and absolute risks, by  $\chi^2$  tests. Length of hospital stay was assessed using non-parametric tests. Tests of significance were two-tailed, and a p value of less than 0.05 or a 95% CI excluding 1 were considered significant. Calculations were performed using Epi Info V.7.1.4.0. We calculated the number of evidence summaries needed to change a care decision for one question raised as the inverse of the risk difference and its CI. In cases where multiple questions related to the care of the same patient were identified, we decided to include all of them in the primary outcome analysis, as we consider that clinician's decisions related to every identified question are independent from one another even when related to the same patient. We also performed a sensitivity analysis in which we considered one question (the first that arose) per patient. In cases where multiple questions related to the care of the same patient were

assigned to different arms, we decided to analyse that patient's clinical outcomes (secondary outcomes) as if assigned to the intervention arm.

#### Results

During 131 morning reports, 553 questions (4.2 questions per meeting) were identified. 398 were excluded because they fulfilled exclusion criteria and 31 were excluded because of lack of time to perform the literature search. In total, 124 questions (62 randomised to each arm), related to the care of 120 inpatients (60 in each group), were included (figure 3) and answered. All the evidence summaries constructed in response to the questions assigned to the intervention arm were effectively delivered to the corresponding treating physicians between 2 and 4 h after their identification. The quality of the evidence found in response to the included questions was judged as very low 32 (26%), low 56 (46%), moderate 32 (25%) and high 4 (3%), and the recommendations were strong for 44 (35%) questions and weak for 80 (65%). Twenty-five (57%) of the strong recommendations were based on moderate or high quality of evidence and 19 (43%) on very low or low quality of evidence (13 based on GRADE paradigmatic situation 1, 2 based on GRADE paradigmatic situation 2 and 4 based on GRADE paradigmatic situation 3)<sup>16</sup> (see online supplementary appendix 2). Baseline question and patients' characteristics were balanced (table 1). Primary outcome (decision/recommendation concordance) occurred in 76 (61.2%) of the included questions, 49 in the intervention arm and 27 in the control arm, relative risk (RR) 1.8 (95% CI 1.3 to 2.4),



**Figure 3** Group assignment and questions identified/answered.

risk difference 35.4% (95% CI 19.5% to 51.4%), number of evidence summaries needed to change a care decision for one question raised 3 (95% CI 2 to 5; [table 2](#)). Sensitivity analysis showed that results were robust if the primary outcome was calculated considering only the first question related to every patient (excluding four extra questions). Subgroup analysis showed that provision of evidence summaries for strong and weak recommendations both have significant impact on recommendation compliance, RR 1.9 (95% CI 1.1 to 2.6) and 1.7 (95% CI 1.1 to 2.6), respectively ([table 3](#)). No significant differences were observed in any of the evaluated secondary outcomes ([table 2](#)).

### Discussion

Ubbink *et al*<sup>18</sup> describe that “evidence-based practice (EBP) provides a structure for the bedside use of research and consideration of patient values, and preferences to optimise clinical decision-making and to improve patient care”. Although the potential benefits of EBP are widely accepted, research consistently shows that there is an important gap between evidence and practice.<sup>19</sup> Some of the identified barriers responsible for this gap include lack of time, resources and/or training.<sup>18</sup> Multiple interventions designed to overcome this gap have been tested with different success rates.<sup>20</sup> These interventions can be classified in three groups of complementary approaches: interventions focused in training clinicians to independently find, appraise and apply the best evidence (ie, training programmes, courses); interventions focused on providing the best

available evidence at the point of care in a friendly and usable format (ie, producing, disseminating and updating trustworthy guidelines); and interventions that directly address behaviour change (ie, reminders, computerised decision support).<sup>21</sup> In this trial, we evaluated an approach that combines the latter two by providing tailored evidence summaries in response to questions that arose during the care of inpatients. The results show that such an approach was able to significantly and frequently influence physicians’ decisions related to their patient care.

Based on the observed results compared with other behavioural changing strategies, the proposed intervention could be categorised as highly effective.<sup>7 22–24</sup> Although we did not observe a significant impact on clinically important outcomes, this was expected in the context of an intervention intended to improve quality of care through affecting physician behaviour, hence improving evidence-based decision-making.<sup>25</sup> Impact on clinically important outcomes could be very difficult to prove (huge sample sizes needed, low signal-to-noise ratio) in this context.<sup>26 27</sup> Thus, process measures such as physician behaviour are considered valid surrogate measures when there is solid evidence of their relation to those clinical important outcomes.<sup>28</sup> Although the intervention in our study included a wide range of recommendations, most of which were based on low-quality evidence, meaning that their impact on clinical important outcomes is dubious, one could argue that improving evidence-based decision-making is a measure of quality by itself.<sup>29</sup> Hence if the proposed intervention

**Table 1** Baseline characteristics

	Intervention	Control	p Value*
<b>Questions</b>	<b>n=62</b>	<b>n=62</b>	
Treatment intervention	44	46	0.6
Answered with strong recommendation	22	22	1
Answer based on moderate or high quality of evidence	21	15	0.2
Strong recommendation based on very low or low quality of evidence	7	12	0.2
<b>Patient who motivated the question</b>	<b>n=60</b>	<b>n=60</b>	
Age† (mean, SD)	64 (17.5)	63.3 (20.4)	0.8
Female (n, %)	32	35	0.5
Severity of the condition that motivated the hospitalisation as subjectively defined by the investigators			
High	20	19	0.8
Low/moderate	40	41	
Charlson comorbidity index	2.8 (2.4)	2.2 (2.4)	0.1
Five most frequent diagnoses			
Cancer complication	6	5	1
Venous thromboembolic disease	3	6	0.4
Stroke	5	3	0.7
Gastrointestinal haemorrhage	2	5	0.4
Pneumonia	4	0	0.6

\* $\chi^2$ .

†Range 18–99.

were able to provide trustworthy recommendations based on the best available evidence, the observed results could be interpreted as a valid surrogate to quality of care improvement. Unfortunately, this cannot be asserted because the evaluated system has not been validated yet, which we assume as the main weakness of our study.

Subgroup analysis showed that strong and weak recommendations both had significant impact on clinicians' decisions, suggesting that they used the provided information as much in the scenario in which the observed results showed that the intervention produced clear benefits or harms, as when they suggested that benefits and risks were balanced, or when there was low/very low quality of evidence or uncertainty about values and preferences. These findings are consistent with other RCTs in which physicians' hypothetical course of action (evaluated with surveys) was influenced by strong as well as weak recommendations.<sup>30 31</sup>

Other studies have evaluated interventions based on evidence facilitating services to improve quality of care;

most of them have observational designs and used surveys to measure their impact on outcomes.<sup>32–35</sup> Their results are inconsistent but suggest possible benefits. Two RCTs informed benefits in terms of physician reported attitude towards searching for information and satisfaction.<sup>9 10</sup> An RCT suggested positive impact through evaluating physicians reporting changes in their behaviour.<sup>36</sup> Another RCT, which evaluated the impact of inserting an evidence statement into hospital discharge letters, showed a significant increase in general practitioner adherence to discharge medications.<sup>11</sup> Our study proposes some novel aspects that could be seen as strengths in the light of the existing evidence. The intervention proposed is based on an evidence analysis system (GRADE approach) that is widely accepted, and has proven to be reproducible and more reliable than intuitive judgement.<sup>37</sup> We included recommendations in the evidence summaries, which have shown to be valuable for clinicians, especially in the context of low or very low quality of evidence, and possibly influence their course of action.<sup>31</sup> The RCT design provides the

**Table 2** Results

	Intervention (n)	Control (n)	RR (95% CI)	p Value	Risk difference (95% CI), %
<b>Primary outcome</b>	<b>n=62</b>	<b>n=62</b>			
Recommendation compliance	49	27	1.8 (1.3 to 2.4)	<0.01	35.4 (19.5 to 51.4)
<b>Secondary outcomes</b>	<b>n=60</b>	<b>n=60</b>			
Inpatient mortality	4	5	0.8 (0.2 to 2.8)	1	–1.6 (–7.7 to 11)
6-month mortality	10	7	1.4 (0.5 to 3.5)	0.4	5 (–7.4 to 17.4)
Intensive care unit	8	10	0.8 (0.3 to 1.8)	0.6	–3.3 (–16 to 9.4)
6-month rehospitalisation	32	27	1.1 (0.8 to 1.7)	0.3	8.3 (–9.4 to 26.1)
Days of hospitalisation (median, IQR)	10	8	–	0.2*	–

\*Mann–Whitney U test.

RR, relative risk.



**Table 3** Subgroup analysis

Outcome	Intervention (n)	Control (n)	Relative risk (95% CI)	p Value	Risk difference (95% CI), %
Strong recommendations (n=44)					
Recommendation compliance	19	10	1.9 (1.1 to 3)	0.01	40.9 (15.6 to 66.1)
Weak recommendations (n=80)					
Recommendation compliance	30	17	1.7 (1.1 to 2.6)	<0.01	32.5 (12.1 to 52.8)

Test for interaction (Breslow-Day test for risk ratio) p value 0.8.

optimal strategy to evaluate quality-improving interventions with the lowest risk of bias.<sup>38</sup> We objectively measured physician behaviour by addressing the proportion of decisions that were coherent with the proposed recommendations in each arm instead of using surveys, which are prone to recall bias.

The main aspects that threaten the confidence and applicability of the observed result are: (1) the reported findings could significantly differ if the proposed interventions were implemented in different contexts (eg, physicians not trained in evidence-based decision-making could be more reticent to use the provided evidence compared with those who participated in the present trial); (2) indirectness of outcome,<sup>39</sup> as mentioned: physician behaviour is a surrogate for important clinical outcomes, the validity of which could be questioned.

### Conclusion

A question identification and answering system was feasible, effectively performed and significantly influenced clinicians' behaviour related to the care of hospitalised patients. The provided information was useful when it resulted in strong as well as in weak recommendations. These findings suggest that interventions that facilitate accessibility and interpretability of the best available evidence at the point of care have the potential to significantly impact on the quality of healthcare.

Twitter Follow Ariel Izcovich at @IzcovichA

Competing interests None.

Ethics approval The German Hospital ethics committee.

Provenance and peer review Not commissioned; externally peer reviewed.

### References

- Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. *Health Info Libr J* 2007;24:78–94.
- 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.
- 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.
- Scott I, Heyworth R, Fairweather P. 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.
- 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.
- Grimshaw JM, Shirran L, Thomas R, et al. Changing provider behavior: an overview of systematic reviews of interventions. *Med Care* 2001;39:112–45.
- Grimshaw JM, Thomas RE, MacLennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004;8:iii–iv, 1–72.
- Vandvik PO, Brandt L, Alonso-Coello P, et al. Creating clinical practice guidelines we can trust, use, and share: a new era is imminent. *Chest* 2013;144:381–9.
- 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.
- Marshall JG, Neufeld VR. A randomized trial of librarian educational participation in clinical settings. *J Med Educ* 1981;56:409–16.
- 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-randomized controlled trial. *Qual Saf Health Care* 2007;16:456–61.
- 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 randomized controlled trial. *Evid Based Med* 2011;16:131–5.
- Richardson WS, Wilson MC, Nishikawa J, et al. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995;123:A12–13.
- 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.
- 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.
- 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.
- Guyatt G, Oxman AD, Sultan S, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol* 2013;66:151–7.
- 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.
- 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.
- Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003;362:1225–30.
- 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.
- Smith WR. Evidence for the effectiveness of techniques to change physician behavior. *Chest* 2000;118(2 Suppl):8S–17S.
- Cheung A, Weir M, Mayhew A, et al. Overview of systematic reviews of the effectiveness of reminders in improving healthcare professional behavior. *Syst Rev* 2012;1:36.

24. Bennett JW, Glasziou PP. Computerised reminders and feedback in medication management: a systematic review of randomized controlled trials. *Med J Aust* 2003;178:217–22.
25. Batalden PB, Davidoff F. What is “quality improvement” and how can it transform healthcare? *Qual Saf Health Care* 2007;16:2–3.
26. Glasziou P. What is EBM? *Evid Based Med* 2011;16:129–30.
27. 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.
28. Lilford RJ, Brown CA, Nicholl J. Use of process measures to monitor the quality of clinical practice. *BMJ* 2007;335:648–50.
29. 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.
30. Cuello García CA, Pacheco Alvarado KP, Pérez Gaxiola G. Grading recommendations in clinical practice guidelines: randomised experimental evaluation of four different systems. *Arch Dis Child* 2011;96:723–8.
31. Neumann I, Schunemann H, *et al.* Do clinicians want recommendations? A randomized trial. NCT02006017. [http://www.gin2014.com.au/d/GIN2014,\\_parallelsession4.5, AbstractBook.pdf](http://www.gin2014.com.au/d/GIN2014,_parallelsession4.5, AbstractBook.pdf) (accessed 20 Nov 2014).
32. Weightman AL, Williamson J, *et al.*, Library & Knowledge Development Network (LKDN) Quality and Statistics Group. The value and impact of information provided through library services for patient care: a systematic review. *Health Info Libr J* 2005;22:4–25.
33. Brettle A, Maden-Jenkins M, Anderson L, *et al.* Evaluating clinical librarian services: a systematic review. *Health Info Libr J* 2011;28:3–22.
34. 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.
35. 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.
36. 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.
37. Mustafa RA, Santesso N, Brozek J, *et al.* The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *J Clin Epidemiol* 2013;66:736–42.
38. Fan E, Laupacis A, Pronovost PJ, *et al.* How to use an article about quality improvement. *JAMA* 2010;304:2279–87.
39. Guyatt GH, Oxman AD, Kunz R, *et al.* GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol* 2011;64:1303–10.