Guidelines helping clinicians make the GRADE

Zubair Bayat · Eddy Lang

Received: 5 July 2014 / Accepted: 29 August 2014 / Published online: 10 September 2014
© SIMI 2014

Introduction

In order to deliver the highest standards of evidence-based care, physicians develop clinical questions that are informed by the literature including recommendations developed by organizations that have the necessary expertise and authority in their respective fields. Guidelines, defined by the IOM as “statements that include recommendations intended to optimize patient care” [1], are a valuable resource increasingly available to clinicians allowing them to make decisions consistent with the best evidence available, while considering the overall balance of benefits and harms as well as resource implications [2]. This paper will explore some of the challenges that have prevented guidelines from being readily interpreted, and hence not being used to their full potential. We will also explain how the GRADE (Grading of Recommendations Assessment, Development and Evaluation) methodology addresses some of the existing challenges, and where it fits in the overall guideline development process. The implications of strong and weak recommendations, as defined by GRADE, on practice by physicians will be discussed as well. Dr. Jake, a fictional emergency medicine physician, will appear periodically throughout the paper to add clinical context and meaning to the discussion.

An emergency department patient encounter

Dr. Jake is an emergency physician working at an urban hospital in a Western Canadian city. At 9:35 AM, Jake assesses a 4-year-old boy brought in by his mother with a 2-day history of non-productive cough. The patient was brought to the emergency department (ED) today as he complained of feeling ill before preschool, and had a measured temperature of 37.9° at home. According to the patient's mother, he had a sore throat for 2 days before the cough started, but it has since resolved. The patient's mother states that the cough is present all day, but she is unsure if it is present at night. The cough is not provoked by positional changes or exertion, and the patient denies shortness of breath, chest pain, having been woken up or kept awake by his cough, and denies experiencing headaches or chills. He also denies any other associated symptoms, and says that since his throat stopped hurting he can eat and drink again. The patient attends preschool, but has been recuperating at home for the past 3 days, (he was scheduled to return today). According to his mother, the patient has no medical conditions, has never been hospitalized, is up to date with his immunizations, and has no known allergies. Cough syrup has not relieved his symptoms, and the patient takes no other medicines. The patient is an only child and lives with both his parents (neither of whom of whom smoke). He has not left Western Canada in his life, and has not recently been exposed to any pets. According to his mother, the patient has never been this sick in the past, and he has been getting progressively more lethargic and less like himself as the illness progresses. She is worried that the patient might have pneumonia or croup, but states that she would prefer that the patient not be prescribed antibiotics if possible, as the patient's family strives to consume organic food exclusively.

Z. Bayat (✉)
Faculty of Medicine, Health Sciences Centre,
University of Calgary Medical School, 3330 Hospital Drive NW,
Calgary, AB T2N 4N1, Canada
e-mail: zbayat@ucalgary.ca

E. Lang
Department of Emergency Medicine, University of Calgary,
Calgary, Canada
e-mail: Eddy.lang@albertahealthservices.ca

The patient appears tired, but is fully alert, oriented and responsive. He is slightly febrile at 37.6°, and normotensive but tachycardic at 120 beats/min, and tachypneic at 34 respirations/min but with no increased work of breathing, and the patient is not in respiratory distress or hypoxemic (his SpO₂ is 99 %.) Dr. Jake hears distinct crackles at the left base, but no wheezes and generally equal air entry in all lung fields is heard. There are no areas of dullness to percussion on the patient's thorax. The rest of the patient's physical examination is normal.

Dr. Jake excuses himself to answer a page, but before returning to see the patient, he contemplates the details of the case, and decides that the most likely diagnosis is community acquired pneumonia (CAP.) Based on the patient's history and physical examination, Dr. Jake is reasonably sure that his patient has a viral pneumonia. Jake remembers learning, as a resident, that bacterial and viral pneumonias cannot be easily distinguished from one another, and that all incidences of CAP in the pediatric population should receive antimicrobial treatment. Dr. Jake's practice has been to do just that, but, in light of the patient's mother's desire to avoid antibiotics, he decides to perform a quick literature search for guidelines that would inform his clinical decision-making. He focuses his search on the management of uncomplicated CAP in the pediatric population. The intervention in question is the administration of antimicrobial therapy, and the comparison, is withholding antibiotic therapy. Dr. Jake decides on effective treatment (the absence of complications and mortality), as his outcome of interest.

Barriers to widespread guideline uptake

Studies have existed endorsing the use of guidelines in many facets of clinical practice for more than a decade [3]. It is perplexing, then, that the uptake of this guidance by clinicians has been inconsistent at best [4]. Unfortunately, this is hardly an isolated case. There are many challenges that have prevented the widespread uptake of guidelines by physicians. While numerous systems are available to help clinicians evaluate the quality of evidence, and guidelines are intended to relieve this burden upon physicians, there are often many competing guidelines pertaining to the same health issues [2]. This abundance of choice can lead to confusion in healthcare providers who are forced to sift through many guidelines to determine which is the most applicable to the clinical circumstances at hand [5]. Furthermore, even when answers to focused questions can be reliably found, there are often difficulties integrating these answers into the broader clinical context [2]. Practitioners must also determine whether or not the patients they are treating are sufficiently similar to the populations for which

the guideline was developed, whether the impact on the patient will be favorable, and whether challenges arising from the reality on the ground, logistics of healthcare delivery or resource limitations will hamper the implementation of the recommendations [6]. Since it is often impractical for clinicians to assess the effect of these variables thoroughly in real time, guidelines that make these considerations explicit reduce unnecessary duplication of efforts by physicians, who might have otherwise had to reappraise the evidence for themselves. Finally, guidelines that are insufficiently transparent with regard to the development of their recommendations, risk instilling skepticism and confusing the end users, and can lead them to decisions that are based on weak evidence, and hence will not be maximally beneficial [7].

Dr. Jake's literature search

The plethora of challenges facing clinicians who seek to use guidelines in the practice of evidence-based medicine are reflected in Dr. Jake's search of the literature to answer his clinical question. A search of the National Guidelines Clearing House (<http://www.guidelines.gov>) found four guidelines purporting to answer Jake's question. Upon closer examination of the guidelines, Jake finds that only two were generated based on studies of his patient's population, and are relevant to his patient's clinical context, and that those were the only two that addressed issues of resource limitation and the logistics of healthcare delivery. He does notice, however, that only one of the two guidelines available to him separates the quality of evidence from the strength of its recommendations, and makes explicit and transparent how it judged quality and strength. Knowing that he must hurry back to treat his patient, Jake is relieved that he has found two guidelines that develop simplified recommendations that he can choose from, and that fit the PICO criteria for his question, instead of being faced with a large number of competing guidelines and no idea as to how to assess the quality of evidence that they base their recommendations on.

An introduction to GRADE

GRADE (Grading of Recommendations Assessment, Development and Evaluation) is a formalized methodology that can be used to evaluate the quality of evidence, synonymous with confidence in estimates of effect in systematic reviews and other evidence bases. It is also used to develop guideline recommendations in a systematic and transparent manner, addressing many of the deficiencies in existing methods of guideline development and evidence

appraisal [7]. GRADE was developed by the GRADE Working Group (GWG), a team of researchers and clinicians at leading evidence-based centers, [7] and has been adopted by many of the premier evidence-based medicine institutes and agencies across the world including the Cochrane Collaboration, the World Health Organization, England's National Institute for Health and Clinical Excellence, (NICE) among many others [8]. The most complete exploration of the benefits and logistics of GRADE methodology can be found in a series on GRADE published in the *Journal of Clinical Epidemiology*, [9] and on the website of the GWG [10].

How GRADE works

When using GRADE methodology to develop guidelines, the clinical questions that are the focus of the guideline project are defined in terms of the population to be studied, the intervention and comparator in question, and all patient important outcomes [11]. The outcomes are then classified as critical and non-critical for decision-making, and guideline authors proceed to define, generally from systematic reviews, the “best estimate of effect” of the intervention on each outcome across studies along with a measure of precision [11]. The evidence is then summarized for each patient important outcome and presented together in an Evidence Presentation (EP) table or Summary of Findings (SoF) table [12]. EP tables contain detailed information about how the guideline developers graded available evidence for a given outcome, and are useful in ensuring that recommendations are made in a systematic and transparent manner, whereas SoFs present the final recommendations in a simplified format, and can be more useful for target users [11].

The quality of evidence is a measure of the confidence that is considered along with the size and direction of an effect [13], and it is a key component in determining the strength of recommendations made based on that evidence [7]. GRADE simplifies the classification of evidence quality/confidence in estimates of effect by assigning a body of evidence for a given outcome to high, moderate, low or very low categories. Failure to recognize low or very low quality evidence can lead clinicians to inadvertently harm patients, as was seen in the misguided use of hormone replacement therapy for the primary prevention of heart disease in healthy post-menopausal women [14]. Well conducted randomized controlled trials (RCT) typically start as high-quality evidence, and observational studies start as low-quality evidence, and a specific set of criteria lead to downgrading, or, more rarely, upgrading a given evidence base. Quality of the evidence is increased for observational studies if the effect size is large, there is a

dose–response gradient, and all possible confounding would reduce a demonstrated effect, or all possible confounding would suggest a spurious effect when the results show no effect [12]. Quality of evidence is, on the other hand, downgraded for both RCTs and non-RCTs if there are serious risks of bias, inconsistency between studies, indirectness, imprecision or suspected publication bias [12] (Table 1).

After adjustments, the quality of evidence for each outcome across studies is ranked from high to very low [11]. Guideline developers review all available information at this point, and make final determinations on the totality of evidence based upon which outcomes are critical for decision-making and, hence, form the basis of a recommendation.

When it comes to developing the actual recommendation, it is not just evidence that is taken into consideration, but the balance between desirable and undesirable outcomes is assessed, along with patient preferences, and costs to the patient and the system. These elements, taken together, determine the direction of the recommendation that guideline developers make [12]. These factors, coupled with the quality of the evidence, inform both the direction and the strength of the recommendations made by the guideline panel [15].

GRADE addresses deficiencies of existing methodologies

Clinicians on the frontlines of delivering patient care are unable to master the intricacies of every guideline development system, and the fact that GRADE has been adopted

Table 1 Summary of reasons to downgrade the quality of evidence with illustrative examples

Reasons to downgrade quality of evidence	Example
High likelihood of bias	Studies conducted with inappropriate controls, many subjects lost to follow-up, or selective reporting of results
Indirectness of evidence	Using outcomes identified in an adult population to make decisions in the pediatric population
Inconsistency of results	Body of evidence consists of studies with inconsistent results
Imprecision of results	Confidence intervals sufficiently wide to cast doubt on precision of results
High likelihood of publication bias	Conclusions drawn from a number of small studies, possibly funded or conducted by parties with conflicts of interest

Adapted from the Cochrane Handbook [19], and the GRADE series in the *Journal of Clinical Epidemiology* [9]

Table 2 Summary of advantages of GRADE over other guideline development modalities, where QoE represents quality of evidence, and SoR represents strength of recommendations

GRADE	Other methodologies
Adopted by leading evidence-based medicine institutions	Not as widely adopted
Transparent and explicit documentation of all steps in guideline development process	Facets of guideline development process often opaque
Overall QoE based on critical outcome with lowest QoE	Determinants of QoE not always explicit
Separates QoE from SoR	QoE and SoR often intertwined in recommendations
Emphasizes the roles of clinical judgement, patient preference and resource considerations in acting on weak recommendations	Significance of weak/conditional recommendations for physician practice often not elucidated
Associated software and programs widely available, free of charge	Associated software often unavailable to public or nonexistent
Useful for systematic reviews, clinical practice guidelines, public health policy, and general evidence assessment	Not usually intended for use outside of systematic reviews and guideline development

Adapted from Guyatt, Oxman et al. [7]

by many leading guideline-producing organizations offers the hope of a common language, which suggests the credibility to serve as the first-line tool for developing recommendations geared towards clinicians [7]. It provides a structured and transparent approach to identifying populations and outcomes of interest, evaluating the evidence in terms of the factors that increase and decrease its quality, upgrading and downgrading the quality of evidence and developing recommendations that consider the practicality of interventions and their concordance with patient preferences and values [12]. GRADE, by basing the judgment of overall quality of evidence on the critical outcome with the poorest quality of evidence, also reduces the chances of assigning undue significance to evidence of inferior quality. GRADE is especially valuable in its clear separation of decisions regarding quality of evidence from strength of recommendations, thereby preventing clinicians from becoming confused and equating high-quality evidence with strong recommendations and vice versa [7]. GRADE also requires the guideline developer to be transparent in the judgment of each factor that influences the quality of evidence by producing EPs that can be relied upon to ensure that the recommendations are made in a systematic and transparent manner [11]. Furthermore, the GWG has made support software that facilitates the implementation of the GRADE methodology available at no charge on its website [16]. GRADE addresses deficiencies in other

methodologies, and stands out as a rigorous, transparent, widely adopted and user-friendly guideline development methodology worthy of being studied and embraced by clinicians (Table 2).

What clinicians must know about GRADE

In order to derive maximum benefit from guidelines, clinicians must also be aware of the recommendations that guidelines can make, and the significance of those recommendations to clinicians' actions. Guidelines produced using the GRADE methodology result in recommendations that are strongly for (benefits clearly outweigh the harms), weakly for (developers less confident in positive balance of consequences), weakly against or strongly against an intervention [13]. To avoid target users conflating weak recommendations with weak evidence, the word "weak" is sometimes replaced with conditional or discretionary [13]. In order to minimize confusion, in addition to stating that a recommendation is either strong or weak, the number "1" is sometimes used to signify a strong recommendation, and the number "2" is used to signify a weak recommendation, though pictorial systems may be employed instead [13]. As per GRADE methodology, the guideline developers make a strong recommendation when they believe that all or almost all informed patients would accept the recommended choice for or against an intervention, whereas the guideline developers make a weak recommendation when they believe that most informed people would make the recommended choice, but a substantial number would not [13]. When weak recommendations are made, the clinician is obliged to spend extra time assessing the specifics of the patient's case, discussing the costs and benefits of the intervention with the patient, determining the patient's perspective and attitude toward the intervention, and involving the patient in the decision on whether or not the intervention is warranted. All of the above steps are equally valuable when guidelines make strong recommendations, but clinicians are advised to use extra time to overcome barriers to implementation or compliance [13]. Statements about the priority of a recommendation do not impact its strength, and are often targeted at makers of health policy as opposed to clinicians delivering patient care [13]. Bearing these facts in mind, clinicians will be able to use GRADE methodology-based guidelines optimally for the benefit of their patients.

Real-time guideline use in the ED

Looking at the guidelines available to him, Dr. Jake notes that both guidelines are applicable to his population of

Table 3 Recommendations and strength of recommendations as detailed in the guidelines published by the British Thoracic Society [17] (BTS) and the Infectious Disease Society of America [18] (IDSA)

	BTS		IDSA	
	Recommendation	Strength	Recommendation	Strength
Antimicrobial therapy	All children with a clear clinical diagnosis of pneumonia should receive antibiotics as bacterial and viral pneumonia cannot reliably be distinguished from each other	[C]—Evidence level IVa (formal combination of expert views)	Antimicrobial therapy is not routinely required for preschool-aged children with CAP, because viral pathogens are responsible for the great majority of clinical disease	(Strong recommendation; high-quality evidence)
CXR	Chest radiography should not be considered a routine investigation in children thought to have community acquired pneumonia (CAP)	[A]—Evidence level Ib (one or more rigorous studies designed to answer the question, but not formally combined)	Routine chest radiographs are not necessary for the confirmation of suspected CAP in patients well enough to be treated in the outpatient setting (after evaluation in the office, clinic, or ED)	(Strong recommendation; high-quality evidence)

interest, and both have guidelines for managing children with suspected CAP in the outpatient population. Neither guideline recommends routine or confirmatory chest radiography (CXR) in children suspected of having CAP (regardless of etiology), so Jake feels comfortable studying the recommendations made about the management of CAP in children. Dr. Jake looks at the British Thoracic Society guidelines, and notices that they recommend a course of antibiotics for all children with a clear clinical diagnosis of pneumonia, stating that bacterial and viral pneumonias cannot be differentiated reliably. [17] The only available information available to Jake about the development of that recommendation is that it was a formal combination of expert views [17]. Jake would have loved to hear more from the guideline developers on this issue, but since no further information was readily available, he turned to the Infectious Diseases Society of America (IDSA) guideline, and saw a recommendation that preschool-age children with CAP not receive antibiotics, as viral pathogens cause the vast majority of CAPs in that population [18]. This recommendation is listed as being strong and coming from high-quality evidence, telling Jake that he can be confident in the estimates of effect, and, that the guideline developers believe the good outcomes of this intervention so greatly outweigh the bad ones that the intervention should be recommended for most people most of the time [13] (Table 3).

After carefully reviewing the recommendations made by both guidelines, Jake chose to implement the IDSA recommendation, as it was based on better evidence, and weighed the positive and negative consequences of withholding antibiotics from children with viral pneumonia in a careful and systematic manner that he understood.

Jake returns to his patient's bedside and explains that his patient has a case of CAP. He tells the patient's mother that, although antibiotics play a vital role in treating

bacterial infections, the best evidence shows that they are not indicated in the treatment of mild, uncomplicated CAP in the pediatric population. He advises that the patient should return to the ED if his CAP becomes acutely worse, and to follow-up with the family's physician if the CAP fails to improve over the coming 3 weeks, while continuing supportive care in the interim. Jake's patient is relieved not to have to take any medications, and the patient's mother is glad that antibiotics would not be necessary to treat the CAP. With both the patient and the mother's fears put to rest, Jake continues on with his busy shift. Being a physician who prides himself on practicing evidence-based medicine and on maintaining the currency of his medical knowledge, Jake is grateful for the opportunity to consult the guidelines and to modify his practice based on the best available evidence. Glancing at the list of patients waiting to be seen, Jake notices that, even today, he will have more opportunities to put his newfound knowledge to use. It is in familiarizing themselves with systems like GRADE, and in taking advantage of the vast amounts of knowledge condensed into specific clinical questions, that clinicians committed to practicing evidence-based medicine can practice their noble profession perched atop the shoulders of giants.

Conflict of interest Dr. Eddy Lang is a member of the GRADE Working Group.

References

1. Institute of Medicine (2011) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. The National Academies Press. <http://www.iom.edu/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx>. Accessed 13 August 2014

2. Straus SE, Shepperd S (2011) Challenges in guideline methodology. *J Clin Epidemiol* 64(4):347–348. doi:[10.1016/j.jclinepi.2011.01.003](https://doi.org/10.1016/j.jclinepi.2011.01.003)
3. Dalinka MK, Alazraki N, Berquist TH, Daffner RH, DeSmet AA, el-Khoury GY, Goergen TG, Keats TE, Manaster BJ, Newberg A, Pavlov H, Haralson RH, McCabe JB, Sartoris D (2000) Imaging evaluation of suspected ankle fractures. American College of Radiology. ACR Appropriateness Criteria. *Radiology* 215 Suppl:239–241
4. Holroyd BR, Wilson D, Rowe BH et al (2004) Uptake of validated clinical practice guidelines: experience with implementing the Ottawa Ankle Rules. *Am J Emerg Med* 22:149–155
5. Manuel DG, Kwong K, Tanuseputro P, Lim J, Mustard CA, Anderson GM, Ardal S, Alter DA, Laupacis A (2006) Effectiveness and efficiency of different guidelines on statin treatment for preventing deaths from coronary heart disease: modeling study. *BMJ* 2006(33):1419
6. Lavis JN, Oxman AD, Souza NM, Lewin S, Gruen RL, Fretheim A (2009) SUPPORT tools for evidence-informed health Policy-making (STP) 9: assessing the applicability of the findings of a systematic review. *Health Res Policy Syst* 7(Suppl 1):S9. doi:[10.1186/1478-4505-7-S1-S9](https://doi.org/10.1186/1478-4505-7-S1-S9)
7. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ, GRADE Working Group (2008) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336(7650):924–926
8. Rehfuess EA, Akl EA (2013) Current experience with applying the GRADE approach to public health interventions: an empirical study. *BMC Public Health* 13:9
9. <http://www.jclinepi.com/content/jce-GRADE-Series>. Accessed 13 August 2014
10. <http://www.gradeworkinggroup.org/index.htm>. Accessed 13 August 2014
11. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, deBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ (2011) GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 64(4):383–394
12. Dijkers M (2013) Introducing GRADE: a systematic approach to rating evidence in systematic reviews and to guideline development. *KT Update* 1 (5)
13. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, Nasser M, Meerpohl J, Post PN, Kunz R, Brozek J, Vist G, Rind D, Akl EA, Schünemann HJ (2013) GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol* 66(7):719–725
14. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML et al (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 288:321–333
15. Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Alonso Coello P, Rind D, Montori VM, Brito JP, Norris S, Elbarbary M, Post P, Nasser M, Shukla V, Jaeschke R, Brozek J, Djulbegovic B, Guyatt G (2013) GRADE guidelines: 15. Going from evidence to recommendation—determinants of a recommendation's direction and strength. *J Clin Epidemiol* 66(7):726–735
16. <http://www.gradeworkinggroup.org/toolbox/index.htm>. Accessed 13 August 2014
17. Harris M, Clark J, Coote N, Fletcher P, Harnden A, McKean M, Thomson A, British Thoracic Society Standards of Care Committee (2011) British Thoracic Society guidelines for the management of community acquired pneumonia in children: update 2011. *Thorax* 66(Suppl 2):ii1–ii23
18. Bradley JS, Byington CL, Shah SS, Alverson B, Carter ER, Harrison C, Kaplan SL, Mace SE, McCracken Jr GH, Moore MR, St Peter SD, Stockwell JA, Swanson JT (2011) The Management of community-acquired pneumonia in infants and children older than 3 months of age: clinical practice guidelines by the Pediatric Infectious Diseases Society and the Infectious Diseases Society of America. Oxford University Press on behalf of the Infectious Diseases Society of America. <http://cid.oxfordjournals.org/content/53/7/e25.full.pdf+html>. Accessed 13 August 2014
19. Higgins JPT, Green S (2011) Cochrane handbook for systematic reviews of interventions version 5.1.0. The Cochrane Collaboration. <http://www.cochrane-handbook.org>. Accessed 13 August 2014
20. Schünemann HJ, Wiercioch W, Etzeandia I, Falavigna M, Santesso N, Mustafa R, Ventresca M, Brignardello-Petersen R, Laisaar KT, Kowalski S, Baldeh T, Zhang Y, Reid U, Neumann I, Norris SL, Thornton J, Harbour R, Treweek S, Guyatt G, Alonso-Coello P, Reinap M, Brozek J, Oxman A, Akl EA (2014) Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ* 186(3)