ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 140−150

Journal of Clinical Epidemiology

GRADE SERIES

GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence

Massimo Brunetti[a,*], Ian Shemilt[b], Silvia Pregno[c], Luke Vale[d], Andrew D. Oxman[e], Joanne Lord[f], Jane Sisk[g], Francis Ruiz[h], Suzanne Hill[i], Gordon H. Guyatt[j], Roman Jaeschke[j], Mark Helfand[k], Robin Harbour[l], Marina Davoli[m], Laura Amato[m], Alessandro Liberati[n], Holger J. Schünemann[j]

[a]*Local Health Unit, Azienda USL Modena, Via S. Giovanni del Cantone 43 Modena, Modena 41100, Italy*
[b]*Institute of Public Health, University of Cambridge, Cambridge, UK*
[c]*University of Modena and Reggio Emilia, Modena, Italy*
[d]*Institute of Health and Society, Newcastle University, Newcastle upon Tyne, UK*
[e]*Norwegian Knowledge Centre for the Health Services, Oslo, Norway*
[f]*Brunel University, Uxbridge, UK*
[g]*Centers for Disease Control and Prevention, Atlanta, GA, USA*
[h]*National Institute of Health and Clinical Excellence, London, UK*
[i]*World Health Organization (WHO), Geneva, Switzerland*
[j]*McMaster University, Hamilton, Ontario, Canada*
[k]*Oregon Health & Science University, Portland, OR, USA*
[l]*Scottish Intercollegiate Guidelines Network, Edinburgh, UK*
[m]*Department of Epidemiology, Lazio Regional Health Service, Rome, Italy*
[n]*University of Modena and Reggio Emilia and Agenzia Sanitaria Regionale, Bologna, Italy*

Accepted 5 April 2012; Published online 3 August 2012

Abstract

Objectives: In this article, we describe how to include considerations about resource utilization when making recommendations according to the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach.

Study Design and Settings: We focus on challenges with rating the confidence in effect estimates (quality of evidence) and incorporating resource use into evidence profiles and Summary of Findings (SoF) tables.

Results: GRADE recommends that important differences in resource use between alternative management strategies should be included along with other important outcomes in the evidence profile and SoF table. Key steps in considering resources in making recommendations with GRADE are the identification of items of resource use that may differ between alternative management strategies and that are potentially important to decision makers, finding evidence for the differences in resource use, making judgments regarding confidence in effect estimates using the same criteria used for health outcomes, and valuing the resource use in terms of costs for the specific setting for which recommendations are being made.

Conclusions: With our framework, decision makers will have access to concise summaries of recommendations, including ratings of the quality of economic evidence, and better understand the implications for clinical decision making.

Keywords: GRADE; Economic evaluations; Costs; Quality of evidence; Risk of bias; Health technology assessment

1. Introduction

In previous articles of this series, we described the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach to formulating a structured clinical question and rating the confidence in effect estimates (quality of evidence) for clinical outcomes. In this

The GRADE system has been developed by the GRADE Working Group. The named authors drafted and revised this article. A complete list of contributors to this series can be found on the *JCE* Web site at www.jclinepi.com.

* Corresponding author. Tel.: +39-059-435-111; fax: +39-059-435-377.

E-mail address: m.brunetti@ausl.mo.it (M. Brunetti).

0895-4356/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jclinepi.2012.04.012

What is new?

Key points

- Grading of Recommendations Assessment, Development, and Evaluation (GRADE) offers a transparent and structured process to include resource use in the development of health care recommendations.
- Important differences in resource use should be included along with other important outcomes in evidence profiles and Summary of Findings tables.
- Key steps in considering resource use are the identification of resource use that is potentially important to decision makers, rating the confidence in effect estimates for important effects on resource use, and valuation of resource use in terms of costs for the specific setting for which ecommendations are being made.

article, we highlight economic outcomes of alternative management strategies or interventions and describe how to include evidence on the impacts of interventions on resource use and costs in the GRADE approach. We focus on challenges with rating the confidence in effect estimates and its reporting in evidence profiles and Summary of Findings (SoF) tables.

2. Resource use and economic evaluation

Health care resources include inputs used at any point in a defined treatment management pathway (Box 1). Non-health care resources include all those inputs provided by other service sectors at any point in the treatment pathway, such as social welfare services (e.g., home adaptation, formal social care, housing) or crime and justice services. Patient and informal caregiver resources include all those inputs provided by patients, their families, or caregivers [1].

What resource use measures to include and the importance placed on each measure depends on whose costs are considered important in a given decision context (the analytic perspective). As the magnitude of the resource use and the value of these resources (i.e., their costs) may vary across (and within) countries and over time, resource use should be measured in natural units, such as the length of inpatient hospital stay in days, or the number of outpatient visits.

Unit costs, the value applicable to a single unit of resource use are also likely to vary across (and within) jurisdictions because of factors such as variations in market prices, economies of scale, and over time due to inflation [1,2].

Box 1 Identifying changes in resource use

1. Changes in use of health care resources

- Intervention (e.g., drugs, surgery, counseling, physical therapy)
  - Land, buildings, equipment
  - Human resources/time
  - Consumable supplies
- Laboratory tests
- Examinations
- Emergency transportation
- Emergency visits
- Hospitalisations
- Specialist visits
- Primary care visits
- Home visits and nursing home visits by health care personnel

2. Changes in use of non-health care resources

- Home adaptation
- Special diets
- Transportation to health care facilities
- Social services (e.g., housing, home assistance, occupational training)
- Crime (e.g., theft, fraud, violence, police investigation, court costs)

3. Changes in use of patient and informal caregiver resources

- Visits
- Hospital admissions
- Patient time for self care
- Time of family or other informal caregivers

4. Changes in productivity

- Time off work because of illness, therapy, or caregiving*

Adapted from Luce et al. [16].

*We suggest that changes in productivity should be captured in the value or importance attached to health outcomes and should not be included as items of resource use.

Economic evaluation is defined as the comparative analysis of alternative interventions in terms of both their costs and effects [1]. There are three main types of economic evaluation in health care: cost-effectiveness analysis, cost-utility analysis, and cost-benefit analysis.

They differ primarily in the approach to the valuation of health outcomes: a single natural or clinical measure in cost-effectiveness analysis; a composite measure of quantity and quality of life in cost-utility analysis (e.g., quality-adjusted life years); and units commensurate with those used to value resource use (usually monetary units) in a cost-benefit analysis. Balance sheets are one way of helping decision makers to explicitly consider resource use along with other outcomes when making recommendations. Box 2 summarizes the advantages and disadvantages of using balance sheets.

Economic evaluations may be conducted concurrently within the framework of an empirical study such as a clinical trial or using a decision model that typically uses secondary data collected from several different sources, including (but not limited to) clinical trials. These two approaches are not mutually exclusive and some level of modeling is necessary, for example, to extrapolate from intermediate to final outcomes [3]. Development of a de novo economic evaluation, alongside evidence profiles, may be useful to move from an evidence summary to a recommendation in a specific context.

Box 2 Balance sheets' advantages and disadvantages

Advantages:

- They condense the most important information to allow efficient processing.
- It is a helpful mechanism for organizing thinking, structuring the analysis of evidence, and focusing debate.
- It explicit judgments about resource use in making recommendations, and can explicit considerations concerning equity.
- They provide the ''raw information'' to which decision makers can apply their own judgments about the trade-offs between health benefits, harms, and use of resources.

Disadvantages:

- When there are complicated trade-offs between multiple outcomes, judgments may require a high level of cognitive processing from the guideline panel members or sometimes could remain implicit, or at best qualitatively described.
- The implicit or qualitative nature of the trade-offs means that it is not possible to ensure that hey are consistent across questions or across guidelines.

In collectively funded health care systems, a decision to treat one individual often entails a loss to other individuals: either through diversion of limited health care resources or increased costs for tax or premium payers. It has been argued that those making treatment and coverage decisions should therefore weigh up evidence for resource use, costs, and relative efficiency of interventions alongside (and incorporating) evidence for their beneficial and adverse effects, and this is increasingly reflected in clinical guideline development processes. However, while there is some evidence of a relatively consistent preference in methods guidelines for use of controlled experimental study designs (e.g., randomized controlled trials [RCTs] or meta-analysis of RCTs) to provide unbiased estimates of effects and resource use, decision makers' needs vary and there is more variability in relation to other methodological components, such as the analytic perspective for costs and the approach to valuation of health and other outcomes [4].

3. The GRADE approach

The GRADE recommends that important differences in resource use should be included along with other important outcomes in evidence profiles and SoF tables. Key steps in considering resources in making recommendations with GRADE are as follows:

1. Identify items of resource use that may differ between alternative management strategies and that are potentially important to patients and decision makers;
2. Find evidence for the differences in resource use between the options being compared;
3. Rate the confidence in estimates of effect; and
4. If the evidence profile and SoF table are being developed to inform recommendations in a specific setting, value the resource use in terms of costs for the specific setting for which recommendations are being made.

In the remaining sections of this article, we will address each of these steps using the example of the opioid replacement program (Tables 1 and 2). Key points in considering resource implications using the GRADE approach are summarized in Box 3.

We suggest that GRADE evidence profiles and SoF tables do not include evidence on relative efficiency derived from previously published or unpublished economic evaluations. This is because economic evaluations often make assumptions that differ substantially from those of guideline developers and use evidence on effects and resource use derived from primary research-based sources that are already summarized in the evidence profile. This does not preclude guideline developers from adapting GRADE evidence profiles and SoF tables to include the results of de novo economic models. However, guideline developers should make clear that this represents a departure from the standard GRADE system.

Table 1. Example of resource use evidence profile

	Quality assessment						Summary of resources and costs			
								Resources costs per patient (1999 AU $)		
Studies (follow-up)	**Design**	**Limitations**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other factors**	**No of patients**	**Methadone**	**Buprenorphine**	**Overall quality**
Drugs (6 mo)										
One study (Doran, 2003)[a]	RCT	No	No	Some uncertainty[b]	No	None	405	Resources (mean daily) 57 mg Costs (6 mo) 37 (33 SD)	11 mg 459 (461 SD)	Moderate ⊕⊕⊕○
Other health care costs (6 mo)										
One study (Doran, 2003)[c]	RCT	No	No	Some uncertainty[b]	No	None	405	Resources NA Costs (6 mo) 1,378 (NA)	NA 1,270 (NA)	Moderate ⊕⊕⊕○
Crime costs										
No information available[d]										

Question: Should buprenorphine maintenance flexible doses vs. methadone maintenance flexible doses be used for opioid maintenance treatment?
Patient or population: Opiate dependents.
Setting: Outpatients in United States, Australia, Austria, Switzerland, and UK.
Viewpoint: societal.
Abbreviations: RCT, randomized controlled trial; NA, not available; SD, standard deviation.
[a] Including dispensing fee.
[b] Includes staff time (i.e., face-to-face contact and preparation time), diagnostic procedures, and facility level (supplies, consumables, capital, equipment, ancillary support including administration, management, security, etc.).
[c] The study was conducted within the Australia health system, while the recommendation was global.
[d] This information was provided only by Harris et al. [10], and it was not considered because the risk of bias was considered too large.

Table 2. Example of summary of findings table

Outcomes	Illustrative comparative risks (95% CI)		Relative effect (95% CI)	Nr. of participants (studies)	Quality of the evidence	Comments
	Assumed risk	Corresponding risk				
	Methadone	Buprenorphine				
Clinical outcomes [13]						
Retention in treatment (after 6–48 wk)	63 per 100[a]	52 per 100 (45–60)	RR 0.82 (0.72–0.94)	976 (7)	High ⊕⊕⊕⊕	
Use of opiate during the treatment[b]		The average difference in SDs for the mean number of morphine positive urinalysis in the intervention group was **0.12 lower** (−0.26 to +0.02).		837 (6)	High ⊕⊕⊕⊕	Data based on morphine urinanalysis; only SMD is provided **Interpretation**: little or no difference
Use of cocaine during the treatment[b]		The average difference in SD for the mean number of cocaine positive urinalysis in the intervention group was **0.11 lower** (−0.03 to +0.25).	—	779 (5)	High ⊕⊕⊕⊕	Data based on urinanalysis; SMD is provided **Interpretation**: little or no difference
Use of benzodiazepine during the treatment[b]		The average difference in SD for the mean number of benzodiazepine positive urinalysis in the intervention group was **0.11 lower** (−0.04 to +0.26).	—	669 (4)	High ⊕⊕⊕⊕	Data based on urinanalysis; SMD is provided **Interpretation**: little or no difference
Criminal behavior[b,c]		The average difference in SD of the mean criminal activity score in the intervention group was **0.14 lower** (−0.41 to +0.14).	—	212 (1)	Moderate ⊕⊕⊕○	Criminal activity as measured by self-report. **Interpretation**: little or no difference
Resource use[d]						
Drugs[e]	57 mg daily 37 AU $ every 6 mo	11 mg daily 422 AU $ more per patient every 6 mo		405 (1)	Moderate ⊕⊕⊕○	Drug and dispensing fee
Other health care costs[e]	1,378 AU $ every 6 mo	108 AU $ less per patient every 6 mo		405 (1)	Moderate ⊕⊕⊕○	Staff time, diagnostic and facilities costs

Question: Should buprenorphine maintenance flexible doses vs. methadone maintenance flexible doses be used for opioid maintenance treatment?
Patient or population: Opiate dependants.
Setting: Outpatients in United States, Australia, Austria, Switzerland, and UK.
Intervention: maintenance flexible doses buprenorphine.
Comparison: maintenance flexible doses methadone.
Abbreviations: CI, confidence interval; RR, risk ratio; SD, standard deviation; SMD, standardized mean difference.

[a] Mean control group values.

[b] ''A standardized mean difference was calculated for continuous outcomes (urine results, self-reported heroin use, and criminal activity). The urine data are presented as a continuous outcome measure but are based on data requested directly from authors. This was necessary as urine results in the literature are routinely reported as the percentage of urine samples collected per treatment group that were positive or negative for a given drug (e.g., heroin) across the study period. This ''count data'' is not compatible with the analyzable data fields in RevMan (i.e., continuous, dichotomous, individual patient data). Based on advice provided by Cochrane statisticians, we asked authors to calculate the number of positive urines for each patient in each treatment group and derive a mean number of positive urines with SD, allowing for analysis of urine results as continuous data.''

[c] Criminal activity measured on a scale (Opiate treatment Index) from 0, no criminal activity, to 16, daily criminal activity in all items.

[d] Crime costs ere not presented because of very low quality.

[e] Costs expressed in AU $ (1999).

Box 3 Key points in considering resource implications using the GRADE approach

- Only important or critical resource use should be included in an evidence profile.
- Evidence must be found providing an estimate of the difference in resource use between the intervention and the comparison group.
- Resource use should be presented in natural units (e.g., days in hospital, minutes of clinician time).
- The quality of evidence should be appraised explicitly for each important or critical resource onsequence using the same criteria as for health outcomes.

4. Identifying potentially important resource use

The first step in identifying important resource use is to clearly state the viewpoint (perspective) from which recommendations are being made. One option is to adopt a societal perspective, that is, a broad viewpoint that includes all important health care, non-health care, and patient and informal caregiver resources, regardless of who pays for them (e.g., third-party payers, patients, families) [5]. This has the advantage of ensuring that who pays does not determine whether an item of resource use is included, thus allowing consideration of the impacts of alternative management strategies on the use of resources across all relevant economic sectors. In Table 1, we adopted a societal perspective in a comparison of two opioid maintenance treatments.

Some guideline developers (e.g., NICE) have a remit to limit considerations of resource use (and costs and relative efficiency) to those resources that incur a cost to the health and social care system. Adopting a health care system perspective implies that important health care resources will be considered while non-health care resources and patient and informal caregiver resources may not be considered. However, this does not preclude consideration of broader (non-health) effects of interventions as outcomes in an evidence profile.

In most health care systems, the costs of health care are typically shared by the government, private insurers, employers, and patients and, even within a society, how costs are shared may differ depending on a patient's age (e.g., whether they are younger than or older than 65 years) or situation (e.g., whether the patient is receiving social welfare assistance). Also, when health services are provided there may be an expectation that any consequent resource use or cost savings to other public bodies or private individuals will result in the transfer of funds to ''compensate'' the health care system for costs it incurs in providing such services. These and other factors may influence the items of resource use considered important when adopting a health care system perspective.

To include an item of resource use in an evidence profile or SoF table, evidence must be found that provides an estimate of the difference in resource use resulting from the implementation of the intervention between the intervention and the comparison group. If no evidence is found, we suggest to include a row stating this for this resource.

For each recommendation, only important items of resource use should be included. We suggest doing this in two steps:

1. Consider whether resource use is important (or critical) for making the recommendation.
2. Consider specific items of resource use and their potential impact on different strategies.

It is also necessary to decide in advance on the period of time over which health outcomes and resource use will be considered (i.e., the time horizon). In Table 1 (an example of an evidence profile), the available evidence only includes outcomes up to 1 year. However, guideline developers are likely to be concerned about longer-term outcomes, in which case it may be appropriate to consider either the short-term outcomes as indirect evidence for longer-term outcomes, or to indicate in an evidence profile that no evidence was found for longer-term outcomes. Because the length of follow-up may vary from outcome to outcome, this should be reported whenever relevant for both health outcomes and resource use (Table 1).

Some outcomes, such as hospitalizations or days in hospital, may be considered to be both important to patients and an important component of resource use. For example, an RCT evaluating the effectiveness of a humanized respiratory syncytial virus monoclonal antibody on viral infections in high-risk infants used hospitalization as a primary clinical outcome [6]. This outcome could also be considered in an evidence profile as a component of resource use. Other patient-important outcomes, such as complications of treatment, do not provide direct measures of the impact of the intervention on resource use, but can be regarded as informative proxies for changes in the use of resources. These types of outcomes should also be considered in an evidence profile as a component of resource use.

Guidelines for reporting economic evaluations [7] recommend that quantities of resource use and unit costs are reported separately (and in addition to measures of costs) to facilitate judgments about the applicability of such evidence. Where economic evaluations comply with this reporting standard, it should be possible to extract quantities of resource use in natural units. GRADE recommends that wherever it is possible to extract quantities of resource use from published reports, these data should be presented in evidence profiles and SoF tables in preference to extracting and presenting data on costs.

However, if published economic evaluations report only measures of costs (e.g., drug costs) or aggregated costs

(e.g., health care costs) guideline developers may not find evidence for all items of resource use that are considered important. In circumstances where costs and/or aggregated costs are clearly attributable to a specific item (or items) of resource use that are considered important to include in a SoF table, a pragmatic decision may be taken to include such measures in the evidence profile, alongside measures of other important items of resource use (if available). If available, information on the unit costs applied in cost calculations should be presented in evidence profiles and SoF tables alongside measures of costs.

Costs and unit costs should be converted to the currency appropriate to the relevant country. Such adjustments can be made using exchange rates based on purchasing power parities (PPPs) and inflation factors. Guidance on the use of PPPs and inflation factors for this purpose and a Web-based conversion tool are available [8].

In our example of an evidence profile comparing buprenorphine and methadone for opioid maintenance treatment (Table 1), information about health outcomes—including criminal behavior—came from a systematic review [9]. Two studies provided evidence for the included outcomes [10,11]. In considering the relative importance of outcomes, resource use was considered to be important. In the next step, drugs, other health care resources, and resource use relating to rates and types of crime were identified as important items of resource use. Travel was not considered important and was excluded from the evidence profile. Criminal behavior was considered both as a patient-important outcome and an informative proxy for resource use because although this outcome does not provide a direct measurement (or valuation) of the impact of the intervention on resource use (or costs), it does provide a proxy indicator of potential changes in the use of resources (e.g., resource use and cost savings resulting from reduced incarceration, reduced victim costs, or redeployment of police and other criminal justice system resources). In the evidence profile (Table 1) and in the SoF table (Table 2), one study [10] was rated as not fulfilling inclusion criteria because it did not meet minimum criteria for avoiding risk of bias.

In general, decisions about whether to generate and present pooled estimates of measures of resource use and costs should be governed by the same principles that apply to health and other patient-important outcomes within the GRADE system, particularly those that relate to considerations of inconsistency [12]. Box 4 summarizes further considerations that may be used to determine the conditions in which it may be judged feasible and useful to present pooled estimates for resources and costs in an evidence profile or SoF table. Meta-analysis of resource use or cost data can be conceptualized as an exploratory approach for synthesizing the results of published (or unpublished) economic analyses as part of the systematic review process [13]. Some analysts have suggested this type of approach as inappropriate or unlikely, arguing that the pooled estimate would not be applicable to any specific decision-making context [14,15].

Box 4 Presenting pooled estimates of resource use and costs [13]

- It is important to distinguish between conditions under which it may be considered appropriate to present pooled estimates of resource use and costs, respectively.
- In general, we can consider presenting pooled estimates of resource use, generated using standard meta-analysis techniques, if we are confident that the ''metric'' in question has a common ''meaning'' across studies (e.g., number of days in the hospital).
- The appropriate conditions for presenting pooled estimates of costs are more controversial and should be considered carefully [14].
- Even in circumstances that the analyst has sufficient information and is sufficiently confident to *generate* pooled estimates, it is necessary for the analyst to make prior adjustments for factors likely to lead to geographical and temporal variations in unit costs (e.g., purchasing power, inflation).
- In practice, the range of circumstances in which it is considered feasible and useful to present pooled estimates of costs is likely to be very limited.
- As for other outcomes, the criteria for pooling and investigating heterogeneity should be pecified a priori, if possible.

As with health outcomes, systematic review authors and guideline developers may need to reassess their initial decisions on both the overall importance of resource use and the importance of specific items of resource use after summarizing the available evidence.

As the choice of appropriate methods to measure and value changes in productivity remains controversial [1], along with others [16], we suggest that changes in productivity should not be included in evidence profiles as a measure of resource use, but only as a health outcome, as time off work because of illness, therapy, or caregiving.

5. Making judgments regarding confidence in estimates of effect for resource use

There are more than 20 published checklists and instruments for assessing the quality of health economic analyses [17]. However, none are specifically constructed to assess the quality of a body of evidence as defined by GRADE—that is, the confidence in estimates of effect [18].

The GRADE recommends that the confidence in effect estimates for each important or critical economic outcome

should be appraised explicitly using the same criteria as for health outcomes. Judgemnts about the confidence in effect estimates should be based, so far as possible, on estimates of resource use, rather than on estimates of the costs of those resources. As with health outcomes, only critical items of resource use should be taken into account in determining the overall confidence across such outcomes.

As for health outcomes, randomized trials start at high quality and observational studies start at low [19]. Observational data can be rated up using the same criteria as for health outcomes [20], and evidence from randomized trials can be rated down [12−21]. It should be noted, for example, that estimates of resource use based on data collected alongside or as part of randomized trials may be unrepresentative of routine practice. This potential threat to external validity occurs because of the influence of ''protocol-prescribed resource use'' (i.e., resource use mandated by the clinical trial design) or ''protocol-derived resource use'' (e.g., increased clinical investigations mandated by trial protocols leading to atypical disease management) [22]. In these circumstances, if the potential threat to the external validity of estimates of resource use is judged to be moderate or high (and assuming the impact cannot be measured to allow it to be factored out of such estimates before rating confidence in effect estimates), guideline developers will rate down the confidence in effect estimates for directness.

As for important health outcomes for which no data are available, lack of information for an important item of resource use should be acknowledged.

5.1. Study limitations (risks of bias)

Risks of bias for estimates of resource use are similar to those for estimates of health outcomes [23]. Nonrandom allocation or inadequate allocation concealment can result in selection bias and important differences, for example, in disease severity, requiring more use of resources. Effective blinding ensures that the compared groups receive a similar amount of attention, ancillary treatment, and diagnostic investigations. Knowledge of intervention assignments may impact on resource use. For example, in a randomized trial [24] comparing early discharge to a hospital at home scheme with continued care in an acute hospital for elderly patients, fewer resources were used by the early discharge group. However, at least some of the difference could be explained by the fact that early discharge patients received care from a resource-focused hospital at home team, which would not be the case in a regular home health care situation.

Incomplete outcome data can bias estimates of resource use. However, if resource use data are missing, but reasons for these are both reported and balanced across groups, the risk of bias is likely to be low. For example, in the study comparing buprenorphine to methadone [11], ''at the patient level, clinical records were reviewed retrospectively for every second patient randomized to each treatment.''

As for health outcomes, adherence to the intention to treat principle is generally necessary to maintain prognostic balance. Investigators violated this principle in the previous study [11]. Patients who entered the study to gain access to buprenorphine but were randomized to methadone either did not commence treatment immediately or withdrew from the study. In either case, they were omitted from the analysis. If patients omitted from the analysis were prognostically different than those included this omission compromised prognostic balance.

Resource use data can be collected directly from patients, in which case there is a risk of recall bias, especially if the recall period is relatively long and detailed information is requested [25]. Validation of self-reported data can reduce the risk of bias. For example, in a study of care for terminally ill patients, data regarding the use of health services was reported by patients and confirmed by providers [26].

It may sometimes be reasonable to assume that there is a high-quality evidence based on assumptions, in particular for use of the intervention. For example in the trial of magnesium sulfate for preeclampsia [27], it is reasonable to assume that patients in the intervention group received magnesium sulfate, whereas those in the comparison group did not.

5.2. Consistency of results

As for health outcomes, consistency of results is likely to be important for resource use. Consistency should be assessed in terms of variations in the magnitude and direction of the difference in resource use across studies. Inconsistencies in results can be expected if there are different patterns of resource use in the settings where studies were conducted, or differences in populations or interventions. When variability exists and investigators fail to identify a plausible explanation, the confidence in the effect estimate decreases. Judgments about the consistency of estimates of resource use can be difficult because of poor reporting of study methods and results, including lack of discussion of study results in the context of the results of previous studies.

5.3. Directness of evidence

Generally, directness of the evidence is likely to be a key consideration in rating the confidence in effect estimates for resource use (and costs). Specifically, it is important to assess the extent to which the available evidence reflects levels and combinations of resource use that are applicable to the setting and population in which the guideline is being developed. As noted above, features of the intervention context may substantially influence the levels and particular combinations of resources needed to provide interventions in different health and social systems and/or service settings. Similarly, it is important to assess whether unit costs underlying estimates of costs are applicable to the decision

makers' setting and, if not, whether they can be adjusted to the target setting to allow the re-estimation of costs.

Ideally, there should be comparable resource use data for an adequate follow-up period for the groups being compared. As discussed above, however, sometimes resource use is not measured for the entire time horizon deemed relevant, but extrapolated from more time-limited measurements. For example, in the trial of antiepileptic drugs for partial epilepsy [27], resources were not collected for the entire 2-year follow-up period and extrapolations were made. "To estimate year 1 costs, the costs in months 10−12 were multiplied by three and added to the costs for months 1−3. Year 2 costs were estimated by multiplying the cost in months 22−24 by four." As with sampling patients, sampling time periods will introduce a risk of bias unless there is a reason to believe that resource use will stable between sampled time points (e.g., for long-term chronic diseases). In this example, this might occur as the quantities of resources used during a specific follow-up time period may be influenced by how long after the initial intervention resource use was measured. For example, the resource use in months 10−12 might have underestimated resources used earlier and overestimated resource use occurring later.

Evidence for resource use and costs from older studies may be indirect because of changes in the use of technologies or innovations in the organization and delivery of care. For example, changes in the way care is delivered because of increased experience (e.g., decreased operating times or systematic reductions in length of stay) or decreasing prices for generic drugs could change prescribing patterns. Indirectness of evidence of costs may also result from differences in providers. For example, teaching and research-based hospitals have higher costs relative to nonteaching hospitals [28].

As a consequence of variations in patterns of resource use (and costs) across settings, guideline developers will frequently choose to focus on the evidence for resource use (and costs) that is most direct, rather than on an average estimate of differences in resource use (and costs) based on pooled evidence derived from studies conducted in several different settings.

5.4. Imprecision

Because of variability in resource use between patients (e.g., some patients use exceptional amounts of costly services), larger sample sizes may be required to ensure that studies are adequately powered to detect differences in resource use between treatment groups compared with health outcomes [29]. Thus, clinical trials may be underpowered to detect differences in resource use [30]. Moreover, studies may not always report confidence intervals or *P*-values for economic estimates [31,32].

5.5. Publication bias

Lastly, as for clinical outcomes [33] economic evaluations are at risk of publication bias [13].

6. Attaching monetary values to resource use

When a recommendation is made in a specific context, attaching appropriate monetary values to quantities of resource use can aid consistent and appropriate valuation of these outcomes by decision makers. In principle, the values should reflect opportunity costs.

So far as possible, monetary valuation of resource use should be made by applying up-to-date and locally relevant unit costs (i.e., applicable to the context of the guideline) to the measured quantity (i.e., number of units) of each item of resource use. Analysis of reliable administrative databases or published data sources for the same jurisdiction are proposed as the most reliable source of data on unit costs [34].

However, if these preferred sources are not available, it may be necessary to use unit costs obtained from previously published studies or other sources. As discussed above, these may need to be adjusted for differences in currency and price year.

Discounting is used in economic evaluations to adjust for social or individual preferences over the timing of costs and health benefits [1]. This means that less weight is given to costs or benefits that occur further in the future than to those expected imminently. For example, in a trial of antiepileptic drugs for partial epilepsy [35], presenting a follow-up of 2 years, the authors discounted costs incurred in the second year. Recommended discount rates differ between countries, and are often varied in sensitivity analyses. It is possible to explicitly discount costs and health outcomes in evidence profiles but this would limit the applicability of the evidence profile to contexts using the same discount rates. We therefore recommend that costs and health outcomes should be reported in evidence profiles in their undiscounted form. However, when costs are presented, these should be reported using the appropriate discount rate for the recommendation context (because costs are already context specific). The data used to calculate these discounted costs—including quantities of all resource items, unit costs, and the discount rate—should be reported. This will enable guideline developers to adapt the cost estimates for their locality.

Similar to other outcomes, it may be appropriate to aggregate different items of resource use. This can be achieved by summing the costs of all included items of resource use, once adjustments for currency and/or price year have been made.

7. Resource use and costs in SoF tables

Table 2 represents a SoF table for the comparison of buprenorphine and methadone for opioid maintenance treatment summarizing the effect estimates and the confidence in those estimates, including resource use and costs. The availability of the evidence profile makes all of the evidence considered for inclusion in the SoF table available to those who want it. In our example, there was little or no difference in health outcomes between buprenorphine and methadone, and buprenorphine cost more. For interventions

that both cost more and are more effective, a SoF table, such as Table 2, does not provide any guidance on whether the net health benefits are worth the additional costs. Such trade-offs must either be made implicitly based on the value judgments of guideline developers, or explicitly based on the outputs of a de novo economic evaluation.

8. Finding economic evidence

Evidence for resource use may be found in a range of research-based sources, including clinical trials, observational studies, technology appraisals, and economic evaluations. It may be published concurrently with, or separately from, reports of clinical studies. Methods for locating previously published and unpublished economic evaluations are summarized elsewhere [36]. Evidence for resource use in a specific setting may be also retrieved from national or local databases, such as drug use from prescription databases or hospitalizations from hospital databases.

9. Conclusions

We described the GRADE approach to rating the quality of economic evidence and how the standard GRADE profile can capture both clinical evidence and data on the resource impact of interventions. Guidelines and recommendations have the potential to help decision makers, clinicians, and patients to improve the quality of care, ensuring the best use of limited resources. Although some guideline developers do not consider resource use and cost explicitly, resource use and costs are just other potential outcomes, such as mortality, morbidity, and quality of life, associated with alternative ways of managing patient problems. It is important that guidelines are built on the best available evidence and that guideline panels use systematic and transparent processes to make judgments about their confidence in effect estimates, moving from the evidence to a recommendation, incorporating considerations of how resources are used. Evidence profiles represent a useful tool to include evidence on the impacts of interventions on resource use and costs in recommendations, focusing on challenges with rating quality of evidence. Although not a requirement for use of the GRADE approach, SoF tables provide succinct, accessible, evidence summary on the important health outcomes and resource consequences, and their quality of evidence. To consider all the relevant resources and costs, it is important that guideline developers include the relevant stakeholders and not just clinicians. With our framework, decision makers will have access to concise summaries of recommendations, including ratings of the quality of economic evidence, and better understand the implications for clinical decision making.

Acknowledgments

This article is dedicated to the memory of Alessandro Liberati, a friend and colleague, who will be greatly missed.

The authors would like to thank Silvia Minozzi and Nick Clark for their help in developing the opioid maintenance treatment example and Mirella Longo for useful comments.

References

[1] Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford, UK: Oxford University Press; 2005.

[2] Sculpher M, Pang F, Manca A, et al. Generalisability in economic evaluation studies in healthcare: a review and case studies. Health Technol Assess 2004;8:1−192.

[3] Buxton MJ, Drummond MF, van Hout BA, Prince RL, Sheldon TA, Szucs T, et al. Modelling in economic evaluation: an unavoidable fact of life. Health Econ 1997;6(3):217−27.

[4] Hjelmgren J, Berggren F, Andersson F. Health economic guidelines—similarities, differences and some implications. Value Health 2001;4(3):225−50.

[5] Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, Liberati A, et al. Grading recommendations: incorporating considerations of resources use. BMJ 2008;336:1170−3.

[6] Impact-RSV Study Group. Palivizumab, a humanized respiratory syncitial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high risk infants. Pediatrics 1998;102:531−7.

[7] Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. BMJ 1996;313:275−83.

[8] Shemilt I, Thomas J, Morciano M. A web-based tool for adjusting costs to a specific target currency and price year. Evid Policy 2010; 6(1):51−9.

[9] Mattick RP, Kimber J, Breen C, Davoli M. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. Cochrane Database Syst Rev 2008;(2):CD002207.

[10] Harris A, Gospodarevskaya E, Ritter A. A randomised trial of the cost effectiveness of buprenorphine as an alternative to methadone maintenance treatment for heroin dependence in a primary care setting. Pharmacoeconomics 2005;23(1):77−91.

[11] Doran CM, Shanahan M, Mattick RP, Ali R, White J, Bell J. Buprenorphine versus methadone maintenance: a cost-effectiveness analysis. Drug Alcohol Depend 2003;71:295−302.

[12] Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. rating the quality of evidence—inconsistency. J Clin Epidemiol 2011; 64:1294−302.

[13] Shemilt I, Mugford M, Byford S, Drummond M, Eisenstein E, Knapp M, et al. Chapter 15: incorporating economics evidence. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Chichester, UK: John Wiley & Sons; 2008:449−79.

[14] Anderson R. Systematic reviews of economic evaluations: utility or futility? Health Econ 2010;19:350−64.

[15] Drummond M. Chapter 15: evidence-based decisions and economics: an agenda for research. In: Shemilt I, Mugford M, Vale L, Marsh K, Donaldson C, editors. Evidence-based decisions and economics: health care, social welfare, education and criminal justice. Oxford, UK: Wiley-Blackwell; 2010:179−85.

[16] Luce BR, Manning WG, Siegel JE, Lipscomb J. Estimating costs in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in health and medicine. New York, NY: Oxford University Press; 1996:176−85.

[17] Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: consensus on Health Economic Criteria. Int J Technol Assess Health Care 2005;21:240−5.

[18] Guyatt GH, Oxman AD, Kunz R, et al, for the GRADE Working Group. GRADE: what is "quality of evidence" and why is it important to clinicians? BMJ 2008;336:995−8.

[19] Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3 rating the quality of evidence—introduction. J Clin Epidemiol 2011;64:401−6.
[20] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines 8: rating the quality of evidence—indirectness. J Clin Epidemiol 2011;64:1303−10.
[21] Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. rating up the quality of evidence. J Clin Epidemiol 2011;64:1311−6.
[22] Coyle D, Lee KM. The problem of protocol driven costs in pharmacoeconomic analysis. Pharmacoeconomics 1998;14(4):357−63.
[23] Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. rating the quality of evidence—risk of bias. J Clin Epidemiol 2011;64:407−15.
[24] Coast J, Richards SH, Peters TJ, Gunnell DJ, Darlow MA, Pounsford J. Hospital at home or acute hospital care? A cost minimization analysis. BMJ 1998;316:1802−6.
[25] Petrou S, Murray L, Cooper P, Davidson LL. The accuracy of self-reported healthcare resource utilization in health economic studies. Int J Technol Assess Health Care 2002;18:705−10.
[26] Hughes SL, Cummings J, Weaver F, Manheim L, Braun B, Conrad K. A randomized trial of the cost effectiveness of VA hospital-based home care for the terminally ill. Health Serv Res 1992;26:801−17.
[27] Simon J, Gray A, Duley L on behalf of the Magpie Trial Collaborative Group. Cost-effectiveness of prophylactic 9996 women with pre-eclampsia from 33 countries: economic evaluation of the Magpie Trial. BJOG 2006;113:144−51.
[28] Baltussen R, Ament A, Leidl R. Making cost assessment based on RCTs more useful to decision makers. Health Policy 1996;37:163−83.
[29] Shemilt I, Mugford M, Vale L, Marsh K, Donaldson C, editors. Evidence-based decisions and economics Oxford, UK: Wiley-Blackwell; 2010.
[30] Briggs A. Economic evaluation and clinical trials: size matters. BMJ 2000;321:132−3.
[31] Doshi JA, Henry AG, Polsky D. Analyses of cost data in economic evaluations conducted alongside randomized controlled data. Value Health 2006;9(5):334−40.
[32] Briggs A, Gray AM. Handling uncertainty when performing economic evaluation of healthcare interventions. Health Technol Assess 1999;3:1−134.
[33] Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines 5: rating the quality of evidence—publication bias. J Clin Epidemiol 2011;64:1277−82.
[34] Cooper N, Coyle D, Abrams K, Mugford M, Sutton A. Use of evidence in decision models: an appraisal of health technology assessments in the UK since 1997. J Health Serv Res Policy 2005;10:245−50.
[35] Marson AG, Al-Kharusi AM, Alwaidh M, Appleton R, Baker GA, Chadwick DW, et al. The SANAD study of effectiveness of carbamazepine, gabapentin, lamotrigine, oxcarbazepine, or topiramate for treatment of partial epilepsy: an unblinded randomised controlled trial. Lancet 2007;369:1000−15.
[36] Glanville J, Paisley S. Chapter 7: searching for evidence for cost-effectiveness decisions. In: Shemilt I, Mugford M, Vale L, Marsh K, Donaldson C, editors. Evidence-based decisions and economics. Oxford, UK: Wiley-Blackwell; 2010.