Journal of Clinical Epidemiology 164 (2023) 45–53

Journal of Clinical Epidemiology

GRADE GUIDANCE SERIES

GRADE guidance 38: updated guidance for rating up certainty of evidence due to a dose-response gradient

M. Hassan Murad[a,b,*], Jos Verbeek[c], Lukas Schwingshackl[d], Tommaso Filippini[e,f], Marco Vinceti[e,g], Elie A. Akl[h,i], Rebecca L. Morgan[b,j], Reem A. Mustafa[b,k], Dena Zeraatkar[l], Emily Senerth[b], Renee Street[m], Lifeng Lin[n], Yngve Falck-Ytter[b,j,o], Gordon Guyatt[i], Holger J. Schünemann[i,p], for the GRADE Working Group

[a]*Mayo Clinic Evidence-Based Practice Center, Rochester, MN, USA*
[b]*Evidence Foundation, Cleveland Heights, OH, USA*
[c]*Department of Public and Occupational Health, Academic Medical Centers Amsterdam, University of Amsterdam, Amsterdam, the Netherlands*
[d]*Institute for Evidence in Medicine, Medical Center − University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany*
[e]*Department of Biomedical, Metabolic and Neural Sciences, Environmental, Genetic and Nutritional Epidemiology Research Center, University of Modena and Reggio Emilia, Modena, Italy*
[f]*School of Public Health, University of California Berkeley, Berkeley, CA, USA*
[g]*Department of Epidemiology, Boston University School of Public Health, MA, USA*
[h]*Clinical Research Institute, American University of Beirut, Beirut, Lebanon*
[i]*Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada*
[j]*School of Medicine, Case Western Reserve University, Cleveland, OH, USA*
[k]*Outcomes and Implementation Research Unit, Department of Internal Medicine, University of Kansas Medical Center, Kansas City, KS, USA*
[l]*Department of Anesthesia, McMaster University, Hamilton, Ontario, Canada*
[m]*South African Medical Research Council, Environment & Health Research Unit, South Africa*
[n]*Department of Statistics, University of Arizona Medical Center-South Campus, Tucson, Arizona, USA*
[o]*VA Northeast Ohio Health Care System, Cleveland, OH, USA*
[p]*Department of Biomedical Sciences, Humanitas University, Milano, Italy*

Accepted 24 September 2023; Published online 29 September 2023

Abstract

Introduction: This updated guidance from the Grading of Recommendations Assessment, Development, and Evaluation addresses rating up certainty of evidence due to a dose-response gradient (DRG) observed in synthesis of intervention and exposure studies.

Study Design and Setting: This guidance was developed using iterative discussions and consensus in multiple meetings and was presented to attendees of the Grading of Recommendations Assessment, Development, and Evaluation Working Group meeting for feedback in November 2022 and for final approval in May 2023.

Results: The guidance consists of two steps. The first is to determine whether the DRG is credible. We describe five items for assessing credibility: a) is DRG identified using a proper analytical approach; b) is confounding the cause of the DRG; c) is there serious concern about ecological bias; d) is the DRG consistent across studies; and e) is there indirect evidence supporting the DRG. The first two of these items are the most critical. If the DRG was judged to be credible, then the second step is to apply the DRG domain and consider rating up, but only by one level due to the concern about residual confounding.

Conclusion: Systematic review authors should only rate up certainty in evidence when a DRG is deemed credible.

Keywords: Dose-response gradient; GRADE; Certainty; Systematic reviews; Meta-analysis; Guidelines; Dose-response meta-analysis; Evidence synthesis

1. Introduction

The dose-response gradient (DRG) is one of the nine causality criteria proposed by Bradford Hill [1] and is one of the three Grading of Recommendations Assessment, Development, and Evaluation (GRADE) domains for rating

Funding Statement: No funding was provided for this work.

* Corresponding author. Mayo Clinic, 200 1st Street SW, Rochester, MN 55905, USA. Tel.: +1-507-284-2560; fax: +1-507-284-4251.
E-mail address: Murad.mohammad@mayo.edu (M.H. Murad).

https://doi.org/10.1016/j.jclinepi.2023.09.011
0895-4356/

What is new?

Key findings

- A new guidance for rating up certainty of evidence due to a dose response gradient (DRG) was developed by the GRADE Working Group

What this adds to what was known?

- We describe 5 items for assessing credibility: a) is DRG identified using a proper analytical approach; b) is confounding the cause of the DRG; c) is there serious concern about ecological bias; d) is the DRG consistent across studies; and e) is there indirect evidence supporting the DRG.
- The 2 most critical items focus on establishing DRG through an appropriate analytic approach and making a judgment about residual confounding which can manifest as a DRG

What is the implication and what should change now?

- Certainty of evidence should not be automatically increased when a DRG is observed. A two-step approach is recommended, starting with establishing credibility and then rating up one level if DRG was deemed credible.

up the certainty in a body of evidence derived from non-randomized studies (NRS) [2,3]. Dose-response gradients (DRGs) are often identified when studying the harmful effects of environmental exposures such as the effect of benzene exposure on childhood leukemia [4] and when evaluating public health interventions such as the effect of antismoking campaigns on youth smoking [5]. Dose-response gradients (DRGs) have also been identified in the context of clinical medicine when studying the effects of an exposure such as the DRG between serum thyrotropin level and thyroid cancer [6], or when evaluating the effects of an intervention, such as the DRG between radiotherapy dose and prostate cancer relapse-free survival [7] or the DRG between psychotherapy delivery and mental health outcomes [8]. Previous GRADE guidance about DRG was brief [2,9] and a methodological survey of the use of DRG when rating certainty in nutrition evidence has shown variability of how this domain was applied in systematic reviews [10].

2. Objectives

The objective of this article is to provide updated and comprehensive guidance about applying the GRADE DRG domain for intervention and exposure studies in an evidence synthesis. We also aimed to update our early considerations about the use of the DRG domain in the context of the ROBINS tools [9].

3. Methods

3.1. Overview

We conducted this work as part of the GRADE project group on environmental health through identification of examples, iterative discussions and presentations at GRADE Project and Working Group meetings (October 2021-May 2023). The core group identified several considerations that facilitate making a judgment about the credibility of a DRG and using it to rate up certainty.

3.2. Examples

We searched literature for the use of the GRADE DRG domain and asked members of the GRADE working group to identify examples from primary studies, systematic reviews, and guidelines that could be used to guide the approach to rating up the certainty of evidence.

3.3. Operationalization of guidance

A writing group developed the approach through iterative discussions and refinement in online meetings and virtual communication. The writing group revised the approach using a consensus-based decision-making process. The guidance was then presented to the entire GRADE working group in November 2022 for discussion and was formally approved by all attendees of the GRADE Working Group meeting in May 2023 without dissenting opinions.

4. Results

4.1. Definition and shapes of the dose-response gradient (DRG)

The DRG has also been called dose-response relation or curve, or a biological gradient, and was originally defined

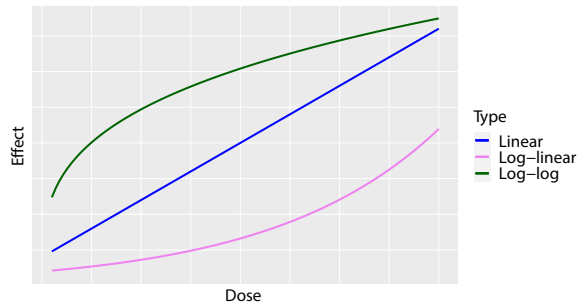

Fig. 1. Monotonic dose-response gradient. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

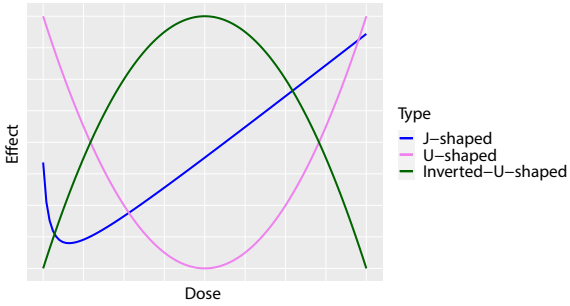

Fig. 2. Nonmonotonic dose-response gradient. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

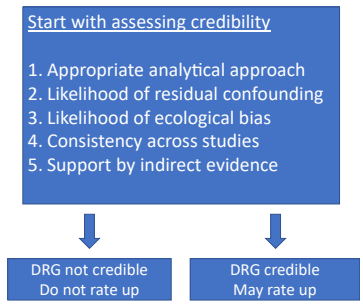

Fig. 3. Items that can be used to assess credibility of a dose-response gradient. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

as a relationship between an exposure and an outcome, in which incremental increases (or decreases) of the exposure produce incremental increases (or decreases) of the outcome [1,11,12]. However, this definition can be expanded to describe other patterns of variation in the outcome in response to variation of the exposure.

DRG shapes are either monotonic or nonmonotonic. A monotonic DRG is a relationship in which the sign of the regression coefficient does not change along the range of the exposure. Thus, if the exposure increases, the response, defined as the occurrence or expression of an outcome or an effect, is always stable or increasing. Monotonic DRGs can be linear, sublinear, supralinear or have a threshold. For many environmental health exposures, DRG is established using a log-linear model, in which the log of the response rate is modeled as a linear function of cumulative exposure [13,14]. Responses with rates that are expected to increase in a linear fashion at lower exposure levels and then plateau or tail off at higher exposure levels are modeled using the "log-log" model, in which the log of the response rate is a linear function of the log of exposure [14]. The log-linear and log-log models have sublinear or supralinear monotonic shapes. Examples of monotonic DRGs are depicted in Fig. 1.

On the other hand, nonmonotonic relationships are the ones in which the regression coefficient sign changes and can have a biphasic shape, a J-shape, U-shape, or inverted U-shape. The concept of hormesis, characterized by stimulation of a response at low dose and inhibition of a response a high dose [15], is a manifestation of a nonmonotonic DRG. Examples of nonmonotonic DRGs are depicted in Fig. 2.

The original epidemiologic papers that discussed DRG as a causality criterion have explicitly described monotonic associations. Bradford Hill used the example of lung cancer mortality increasing linearly with the number of cigarettes smoked daily [1]. Rothman defined the biologic gradient as a unidirectional monotonic dose-response curve [12]. The examples used in the previous GRADE guidance for rating up certainty were all monotonic [2]. The monotonic DRG shape is more intuitive as a causality criterion or a rationale for increasing certainty (*more of the intervention/exposure leads to more of the outcome*). Nonmonotonic DRG, as a criterion to increase certainty of the evidence is more challenging to explain, and therefore, may require more elaboration. The rationale for increasing certainty in the case of a nonmonotonic DRG is not "*more of the exposure leads to more of the outcome*", but rather, reflects that the DRG allows better understanding of the biological relationship between the exposure and outcome. In other words, being able to mathematically predict the outcome based on the exposure. This is particularly more understandable when the nonmonotonic DRG is demonstrated consistently across multiple studies which may serve as a validation of a model in different contexts. For example, a meta-analysis of 250 cohort studies suggested a J-shaped nonmonotonic association between body mass index and all-cause mortality [16]. This DRG was present with some minor variations in shape, in men and women, Whites and African Americans, smokers and nonsmokers; and can be viewed as being validated in different contexts.

4.2. Updated GRADE guidance for rating up certainty due to the dose-response gradient (DRG)

The guidance consists of two steps. The first is to determine whether the DRG is credible. If the DRG was judged to be credible, then the second step is to apply the DRG domain and consider rating up.

4.2.1. The first step: credibility of a dose-response gradient (DRG)

We propose five items that systematic reviews authors can use to make a judgment about the credibility of a DRG (Fig. 3).

4.2.1.1. Is the dose-response gradient (DRG) identified using a proper analytical approach?. Formal statistical testing for DRG in a systematic review can be done using different approaches based on the availability of data. One approach is an analysis of subgroups of studies that evaluated the effect (exposure vs. no exposure) at different levels. Another approach is to conduct a dose-response meta-analysis or metaregression that takes into account all the doses in the included studies. Analysis models can

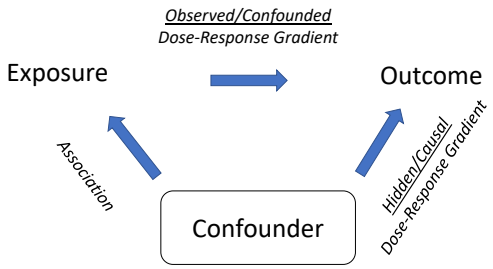

Fig. 4. Depiction of how confounding can manifest as a dose-response gradient. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

be linear such as the generalized least squares for trend, or nonlinear such as the restricted cubic splines function. The rationale for nonlinear DRG has increased in recent years where simple linear models in environmental health, nutrition and toxicology are thought to be too simplistic and may not accurately characterize a DRG [17−19]. A credible model needs to make sense biologically, have a good fit, and be derived from a sufficient number of datapoints.

For example, a meta-analysis evaluating the association of rofecoxib with cardiovascular events suggested a DRG as doses of 25 mg/d or less had a lower relative risk (RR) than doses more than 25 mg/d (1.33 (95% confidence interval, 1.00−1.79) and 2.19 (95% confidence interval 1.64−2.91); respectively) [20]. Despite the statistically significant interaction between the effects of the two doses, this DRG is based only on two doses, and its shape cannot be reliably extrapolated beyond these two doses. More datapoints in the DRG model (i.e., more dose-response pairs) would be needed to reduce the potential for chance findings and provide more certainty for the shape of the DRG. In a simple linear regression model with dose as a predictor variable, a minimum of three dose-response data points would be needed because each of the regression intercept, slope and error term requires a degree of freedom. Thus, a DRG that is based only on two doses is clearly less credible. Evaluating nonlinear and nonmonotonic DRGs, and the use of semiparametric methods, requires even a larger number of observations.

4.2.1.2. Is confounding the cause of the dose-response gradient (DRG)?. While the evaluation of confounding is an established part of the GRADE domain that deals with study limitations, this domain should be revisited when we contemplate rating certainty up due to a DRG. A DRG does not make confounding less likely [21]. In fact, confounding can cause a DRG to manifest between an exposure and an outcome of interest despite the absence of a causal relation. This happens if two conditions are met: a DRG exists between a confounder and the outcome and a close association exists between the confounder and the exposure [12,21,22]. This phenomenon is depicted in Fig. 4.

For example, several case-control studies showed a DRG between coffee consumption and pancreatic cancer [23,24]. The adjusted RRs for consumers of no cups per day, one to two, three to four, and at least five were 1.0, 2.1, 2.8, and 3.2, respectively [23]. This DRG may erroneously lead us to infer a causal relation that, as it turns out, prospective studies have shown does not exist. The cause of this observed DRG between coffee consumption and pancreatic cancer is that the confounder, which is smoking, has a DRG with pancreatic cancer and is closely associated with coffee drinking. A meta-analysis of 42 NRS's found that not smoking, smoking 10, 20, 30, and 40 cigarettes per day was associated with RRs of 1.5, 1.9, 2.0, and 2.1, respectively [25,26]. In addition, smoking and coffee drinking have a strong positive association as each additional cigarette smoked per day is associated with higher coffee consumption of 0.10 cups per day [27,28]. Thus, the "hidden DRG" between the confounder and the outcome, along with close association between the confounder and the exposure, present as a DRG between the exposure and the outcome. Another example is the observed DRG between the incidence of trisomy 21 syndrome and birth order [29]. This observed DRG is caused by a hidden DRG between the confounder, which is the maternal age, and the outcome, along with close association between maternal age and birth order [12].

These two conditions required for a DRG to manifest because of confounding are expected to be common. Since the confounder is a causal factor for the outcome, it is expected to have a DRG with the outcome. And by definition, the confounder is expected to have a close association with the exposure. For example, smoking has an established DRG with various cancers (pancreas, lung, nasopharyngeal, and bladder) [30−32] and may confound the associations of various exposures with these cancers. If smoking and these exposures are closely associated, we would observe a DRG between the exposures and cancer.

Due to the concern about confounding and to avoid erroneously increasing certainty in a confounded association, we recommend that increasing certainty due to DRG requires meticulous evaluation of confounding. Certainty should not be increased due to an observed DRG if the analysis does not adjust for plausible confounders or when known residual confounding remains a serious concern.

There will be cases in which the effect of confounding on DRG is not as clear as the above presented examples. For instance, a DRG exists between red meat and cardiovascular outcomes [33], but red meat exposure tracks closely with many other aspects of dietary habits, which makes determining the causal culprit more challenging. Nevertheless, because of the high suspicion of residual confounding in this association, we do not recommend rating up for a DRG.

4.2.1.3. Is there serious concern about ecological bias?. If a DRG can be demonstrated within studies and between studies, the credibility of the DRG will be higher. When the DRG is only established via between-study

comparisons, it is at increased risk of ecological bias. Ecological bias occurs when inferences about individuals are assumed from inferences about the groups to which they belong [34]. Therefore, to rate certainty up due to DRG, this ecological bias should be deemed unlikely when risk of bias (RoB) is assessed across studies. One example is a meta-analysis that demonstrated a DRG across studies evaluating the association between egg consumption and heart failure, but several included studies in this meta-analysis did not demonstrate a DRG [35]. Such a DRG could be attributed to the ecological bias caused by factors at a study level, such as the geographic location of the studies or their population characteristics.

Nevertheless, the demonstration of a DRG only across studies may be the only approach to achieve sufficient variability in the exposure. A dose-response meta-analysis demonstrated a linear relationship between the stringency of governmental measures to contain COVID-19 and worsening depression and anxiety. Such analysis must depend on comparison of different studies conducted in different locations that have variation in COVID-control measures [36].

4.2.1.4. Is the dose-response gradient (DRG) consistent across studies?. While inconsistency is a separate GRADE domain, when rating up for DRG is being considered, review authors should specifically evaluate whether DRG magnitude and shape is consistent across studies. There are examples of inconsistent conclusions about DRG between studies such as the DRG between potato intake and the risk of type 2 diabetes [37], and consistent ones such as the DRG between physical activity and cardiorespiratory fitness and the incidence of heart failure [38]. Assessing inconsistency of DRG across studies is a matter of judgment because there are no reliable statistical methods to do so. Modeled DRGs of individual studies can be visually depicted which may help in assessing inconsistency, such as the example of the modelled DRGs of cadmium exposure and risk of prediabetes; which visually appeared to be consistent [39]. When evaluating inconsistency, it important to recognize that smaller studies may not show a DRG because they are underpowered. Thus, if a small study did not show a DRG consistent with that of other larger studies, this may not necessarily imply inconsistency.

4.2.1.5. Is there indirect evidence supporting the dose-response gradient (DRG)?. The presence of indirect evidence, such as biological plausibility or mechanistic evidence that corroborates or explains the DRG, can make it more trustworthy. For example, cellular and mechanistic explanations have been proposed for DRGs demonstrated in the associations between lung cancer and inorganic arsenic concentrations [40] and in the association between alcohol intake and colorectal cancer [41]. In contrast, there was no clear biological evidence to support the DRG observed in the inverse associations between sucrose intake and risk of type 2 diabetes [42] or between chocolate intake and the risk of type 2 diabetes [43].

4.2.2. The second step: applying the DRG domain

4.2.2.1. Rating up. If DRG is deemed credible, review authors can consider rating up. Because of the concern about confounding, we recommend not to increase certainty due to DRG by more than one level, even if authors have failed to identify a plausible confounder. In addition, rating certainty up after identifying reasons to rate it down is discouraged in GRADE.

Lastly, it is important to note that some of the information that has been used to judge credibility of the DRG may have been collected during the assessment of other GRADE certainty domains, such as risk of bias and inconsistency, and we should avoid rating down twice for the same concerns.

4.2.2.2. Working with different risk of bias tools. Conventionally in GRADE, the body of evidence from NRS starts at low certainty of evidence due to the concern about residual confounding and selection bias. However, two instruments assess RoB of NRS along an absolute scale (Risk of Bias in Nonrandomized Studies of Intervention and Exposures - ROBINS-I and ROBINS-E) [44]. If these instruments are used, the body of evidence starts at high certainty with the understanding that the it is usually rated down by at least two level [9]. Until release of GRADE's full update of the guidance on rating study limitations and execution of risk of bias across a body of evidence when these instruments are used, we suggest starting with making a judgment about RoB in each individual study using ROBINS-I, then making a global judgment about RoB across the whole body of evidence for each outcome. In this framework, a credible DRG can be viewed as a way to mitigate rating down due to RoB, as long as the rating down was not due to serious concern about residual confounding.

4.2.2.3. Choosing estimates to include in the summary of findings table. Four ways of presenting effect estimates have been noted in the context of rating up for DRG [10]. The first two ways compared any level of the exposure vs. no exposure (e.g., smoking any amount vs. no smoking), or compared extreme categories of the exposure (e.g., highest vs. lowest quantiles of packs per day smoked). Analyses using these two types of estimates are expected to have a high level of heterogeneity due to variation of the effect across doses of the exposure. The heterogeneity would be higher in the first type and may invalidate the homogeneity assumption of a pooled estimate across doses. Comparing extreme exposure categories has poor statistical properties,

can produce larger effects compared to using data about all available doses of exposure [45], and is also susceptible to misclassification error when the exposure is assigned into discrete categories [46]. These two types of estimates are more conducive to a yes/no answer and will be appropriate when the review question has this binary purpose. The third way is to present the effect is per unit of exposure, which is expected to have the least heterogeneity. The fourth way is to present multiple effect estimates for different levels of exposure. This latter approach is most conducive for decision-making and trading off benefits and harms at a certain exposure level.

The choice between these different types of effect estimates depends on the level of contextualization and the goal of the evidence synthesis and subsequent decision [47−50]. For example, let us consider that the question being addressed in evidence synthesis is: *does regular physical activity reduce the risk of severe COVID-19 infection, and is there a threshold of physical activity associated with this effect?* A minimally contextualized approach with a non-null threshold can be pursued to answer this question. A systematic review of NRS demonstrates a nonlinear DRG between physical activity presented in metabolic equivalent of task (MET)-min per week and severe COVID-19 illness [51]. The DRG can increase certainty in a non-null target of certainty concluding that those who engaged in regular physical activity developed less severe disease than inactive peers [51]. The results also showed a flattening of the dose-response curve at around 500 MET-min per week, suggesting that the greatest benefit is attained by performing at least 150 minutes of moderate-intensity aerobic exercise each week [51].

Conversely, let us consider that the question being addressed in evidence synthesis is: *what are the benefits and harms of a specific dose of salvage radiotherapy following radical prostatectomy*? A partially or fully contextualized approach is needed to properly address this question. A dose-response meta-analysis has demonstrated a nonlinear DRG between radiotherapy dose and relapse-free survival with a 2% improvement for each 1 Gy of salvage radiotherapy [7]. The effect estimates to be graded in this question would be the ones estimated at specific doses considered by stakeholders to be most relevant to the clinical context (i.e., radiation doses commonly used in contemporary clinical practice). Authors of this meta-analysis proposed salvage radiotherapy doses of 76 Gy and 66 Gy as important clinical dilemmas and proposed a future trial at these doses [7]. Thus, decision thresholds can be established to support rating certainty at these two levels of exposure. It is also plausible to apply both views as two consecutive steps (i.e., first evaluate the certainty in the exposure itself in a binary fashion, then evaluate certainty in specific estimates at specific exposure levels most relevant to decision-making). Paradigms for framing systematic review questions for exposures have been proposed and can help operationalize decision-making questions [52].

5. Discussion

5.1. Summary

We developed new GRADE guidance to facilitate the application of the DRG as a domain for rating up the certainty of evidence derived from NRS. A key first step is to make a judgment about the credibility of the DRG. The two most critical credibility items focus on establishing DRG through an appropriate analytic approach and making a judgment about residual confounding.

5.2. What does the dose-response gradient (DRG) make us more certain about?

To answer this question, we pose three probing questions. First, does a DRG increase our certainty in a non-null effect? The answer to this question is ‘‘yes’’. Causality is a binary construct that GRADE attempts to address through its leveled certainty ratings. A DRG increases the certainty that an intervention or exposure causes the outcome. The presence of DRG in this view, by definition, implies that the effect estimates vary based on the exposure dose. Thus, there is more certainty about the fact that varying levels of exposure cause varying levels of the outcome, rather than supporting a single specific estimate.

Second, does a DRG increase our certainty in multiple effects, each estimated at a different exposure dose? The answer to this question is ‘‘yes’’. Let us assume that we have five exposure levels with five associated RR’s (RR_1-RR_5) that demonstrate a linear DRG. A DRG in this case increases certainty that at an exposure level between 3 and 4 (e.g., exposure level 3.5), the RR would be between RR_3 and RR_4. Of note, these five levels can be underpinned with different bodies of evidence and there may be cases in which certainty would be lower at the highest and lowest exposure levels than intermediate levels. Because the location of the effect for a specific exposure level is more specified within a particular range, this view allows contextualization and relating the observed estimates to specific decisional thresholds and ranges. In the case of nonmonotonic DRG, the practical application of this view is less clear than for monotonic DRGs. Replication of DRG in multiple studies is important in this context. Rating up certainty in the estimates due to DRG involves circular reasoning because these are the same estimates that generated the DRG. Replication of DRG in multiple studies can alleviate this concern.

Third, does a DRG increase our certainty that the effect is large? The answer to this question is ‘‘no’’. The effects, across a spectrum of exposures, may be large or small despite a DRG. Many environmental exposures with a DRG have a small effect size even for extreme exposures.

Despite the important public health impact of these exposures due to the large part of the population being exposed, the actual effect size or magnitude of association could be quite small (e.g., RR of 1.01 for the effect of air pollution on the risk of myocardial infarction [53]).

We conclude that a DRG increases the certainty about a non-null effect of the exposure, as well as about multiple effects estimated at different exposure levels.

5.3. Strengths and limitations

Compared with the previous GRADE guidance [2], the availability of many applied examples of systematic reviews that invoked the DRG domain and advances in the analytical methods of establishing a DRG have facilitated the development of this guidance. This guidance provides practical steps that prompt systematic review authors to address critical issues such as residual confounding and ecological bias, and not just simply rate up whenever a DRG is observed.

While analytic approaches of DRG have advanced, methodological research about rating certainty up due to DRG remains quite limited. In addition, it remains unclear whether approaches that quantify the importance of residual confounding [54] can aid in determining the impact of confounding on DRG. This guidance suggests that indirect evidence supporting a DRG can make it more trustworthy, however, we recognize that providing a biological rationale to explain observed associations can be misleading and is often misused [22,55]. Lastly, DRG can be identified within and across randomized trials. However, rating up in GRADE and this guidance, focus on NRS.

6. Conclusions

Systematic review authors contemplating rating up the certainty in evidence due to a DRG should first evaluate the credibility of the DRG, with particular attention to the analytic approach that established the DRG and the potential for residual confounding. Concerns about credibility of DRG in these two items should lead to a decision of not rating up. DRG can increase certainty in both, a non-null effect or in multiple effects estimated at different exposure doses.

CRediT authorship contribution statement

M. Hassan Murad: Conceptualization, Methodology, Project administration, Writing − original draft. **Jos Verbeek:** Methodology, Writing − original draft, Writing − review & editing. **Lukas Schwingshackl:** Methodology, Writing − original draft, Writing − review & editing. **Tommaso Filippini:** Methodology, Writing − original draft, Writing − review & editing. **Marco Vinceti:** Methodology, Writing − original draft, Writing − review & editing. **Elie Akl:** Methodology, Writing − original draft, Writing − review & editing. **Rebecca L. Morgan:** Methodology, Writing − original draft, Writing − review & editing. **Reem A. Mustafa:** Methodology, Writing − original draft, Writing − review & editing. **Dena Zeraatkar:** Methodology, Writing − original draft, Writing − review & editing. **Emily Senerth:** Methodology, Writing − original draft, Writing − review & editing. **Renee Street:** Methodology, Writing − original draft, Writing − review & editing. **Lifeng Lin:** Methodology, Writing − original draft, Writing − review & editing. **Yngve Falck-Ytter:** Methodology, Writing − original draft, Writing − review & editing. **Gordon Guyatt:** Conceptualization, Methodology, Writing − original draft, Writing − review & editing. **Holger J. Schünemann:** Conceptualization, Methodology, Writing − original draft, Writing − review & editing.

Data availability

No data were used for the research described in the article.

Declaration of competing interest

Authors of this paper are GRADE Working Group members. Gordon H. Guyatt and Holger J. Schünemann are the co-founders and co-chairs of GRADE Working Group. Gordon H. Guyatt is a member of the Journal of Clinical Epidemiology Editorial Board.

References

[1] Hill AB. The environment and disease: association or causation? Proc R Soc Med 1965;58:295−300.
[2] 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.
[3] Schunemann H, Hill S, Guyatt G, Akl EA, Ahmed F. The GRADE approach and Bradford Hill's criteria for causation. J Epidemiol Community Health 2011;65:392−5.
[4] Filippini T, Hatch EE, Rothman KJ, Heck JE, Park AS, Crippa A, et al. Association between outdoor air pollution and childhood leukemia: a systematic review and dose-response meta-analysis. Environ Health Perspect 2019;127:46002.
[5] Farrelly MC, Davis KC, Haviland ML, Messeri P, Healton CG. Evidence of a dose-response relationship between "truth" antismoking ads and youth smoking prevalence. Am J Public Health 2005;95:425−31.
[6] McLeod DS, Watters KF, Carpenter AD, Ladenson PW, Cooper DS, Ding EL. Thyrotropin and thyroid cancer diagnosis: a systematic review and dose-response meta-analysis. J Clin Endocrinol Metab 2012;97:2682−92.
[7] King CR. The dose-response of salvage radiotherapy following radical prostatectomy: a systematic review and meta-analysis. Radiother Oncol 2016;121:199−203.
[8] Robinson L, Delgadillo J, Kellett S. The dose-response effect in routinely delivered psychological therapies: a systematic review. Psychother Res 2020;30:79−96.

[9] Schunemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. J Clin Epidemiol 2019; 111:105−14.
[10] Eble J, Harms L, Verbeek J, Morgan RL, Schunemann HJ, Meerpohl JJ, et al. The use of the GRADE dose-response gradient domain in nutrition evidence syntheses varies considerably. J Clin Epidemiol 2022;146:12−21.
[11] Shimonovich M, Pearce A, Thomson H, Keyes K, Katikireddi SV. Assessing causality in epidemiology: revisiting Bradford Hill to incorporate developments in causal thinking. Eur J Epidemiol 2021;36:873−87.
[12] Rothman KJ, Greenland S. Hill's Criteria for Causality. In: Balakrishnan N, Colton T, Everitt B, Piegorsch W, Ruggeri F, Teugels JL, editors. Wiley StatsRef: Statistics Reference Online. https://doi.org/10.1002/9781118445112.stat05168.
[13] Saltzman BE. Health risk assessment of fluctuating concentrations using lognormal models. J Air Waste Manag Assoc 1997;47: 1152−60.
[14] Steenland K, Deddens JA. A practical guide to dose-response analyses and risk assessment in occupational epidemiology. Epidemiology 2004;15:63−70.
[15] Calabrese EJ. Hormetic mechanisms. Crit Rev Toxicol 2013;43: 580−606.
[16] Aune D, Sen A, Prasad M, Norat T, Janszky I, Tonstad S, et al. BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74 million deaths among 30.3 million participants. BMJ 2016;353:i2156.
[17] Hill CE, Myers JP, Vandenberg LN. Nonmonotonic dose-response curves occur in dose ranges that are relevant to regulatory decision-making. Dose Response 2018;16:1559325818798282.
[18] Vinceti M, Filippini T, Malavolti M, Naska A, Kasdagli M, Torres D, et al. Dose-response relationships in health risk assessment of nutritional and toxicological factors in foods: development and application of novel biostatistical methods. EFSA Supporting Publication 2020;17:1899E.
[19] Filippini T, Naska A, Kasdagli MI, Torres D, Lopes C, Carvalho C, et al. Potassium intake and blood pressure: a dose-response meta-analysis of randomized controlled trials. J Am Heart Assoc 2020;9: e015719.
[20] McGettigan P, Henry D. Cardiovascular risk and inhibition of cyclooxygenase: a systematic review of the observational studies of selective and nonselective inhibitors of cyclooxygenase 2. JAMA 2006; 296:1633−44.
[21] Rosenbaum PR. Does a dose-response relationship reduce sensitivity to hidden bias? Biostatistics 2003;4:1−10.
[22] Weiss NS. Inferring causal relationships: elaboration of the criterion of "dose-response". Am J Epidemiol 1981;113:487−90.
[23] MacMahon B, Yen S, Trichopoulos D, Warren K, Nardi G. Coffee and cancer of the pancreas. N Engl J Med 1981;304:630−3.
[24] Clavel F, Benhamou E, Auquier A, Tarayre M, Flamant R. Coffee, alcohol, smoking and cancer of the pancreas: a case-control study. Int J Cancer 1989;43:17−21.
[25] Zou L, Zhong R, Shen N, Chen W, Zhu B, Ke J, et al. Non-linear dose-response relationship between cigarette smoking and pancreatic cancer risk: evidence from a meta-analysis of 42 observational studies. Eur J Cancer 2014;50:193−203.
[26] Molina-Montes E, Van Hoogstraten L, Gomez-Rubio P, Lohr M, Sharp L, Molero X, et al. Pancreatic cancer risk in relation to lifetime smoking patterns, tobacco type, and dose-response relationships. Cancer Epidemiol Biomarkers Prev 2020;29:1009−18.
[27] Treur JL, Taylor AE, Ware JJ, McMahon G, Hottenga JJ, Baselmans BM, et al. Associations between smoking and caffeine consumption in two European cohorts. Addiction 2016;111: 1059−68.
[28] Bjorngaard JH, Nordestgaard AT, Taylor AE, Treur JL, Gabrielsen ME, Munafo MR, et al. Heavier smoking increases coffee consumption: findings from a Mendelian randomization analysis. Int J Epidemiol 2017;46:1958−67.
[29] Stark CR, Mantel N. Effects of maternal age and birth order on the risk of mongolism and leukemia. J Natl Cancer Inst 1966;37: 687−98.
[30] Wang L, Mai ZM, Ngan RK, Ng WT, Lin JH, Kwong DL, et al. Dose-response reduction in risk of nasopharyngeal carcinoma from smoking cessation: a multicenter case-control study in Hong Kong, China. Front Oncol 2021;11:699241.
[31] Vineis P, Alavanja M, Garte S. Dose-response relationship in tobacco-related cancers of bladder and lung: a biochemical interpretation. Int J Cancer 2004;108:2−7.
[32] Gandini S, Botteri E, Iodice S, Boniol M, Lowenfels AB, Maisonneuve P, et al. Tobacco smoking and cancer: a meta-analysis. Int J Cancer 2008;122:155−64.
[33] Zeraatkar D, Han MA, Guyatt GH, Vernooij RWM, El Dib R, Cheung K, et al. Red and processed meat consumption and risk for all-cause mortality and cardiometabolic outcomes: a systematic review and meta-analysis of cohort studies. Ann Intern Med 2019; 171:703−10.
[34] Richardson S, Stucker I, Hemon D. Comparison of relative risks obtained in ecological and individual studies: some methodological considerations. Int J Epidemiol 1987;16:111−20.
[35] Bechthold A, Boeing H, Schwedhelm C, Hoffmann G, Knuppel S, Iqbal K, et al. Food groups and risk of coronary heart disease, stroke and heart failure: a systematic review and dose-response meta-analysis of prospective studies. Crit Rev Food Sci Nutr 2019;59: 1071−90.
[36] Salanti G, Peter N, Tonia T, Holloway A, White IR, Darwish L, et al. The impact of the COVID-19 pandemic and associated control measures on the mental health of the general population : a systematic review and dose-response meta-analysis. Ann Intern Med 2022;175: 1560−71.
[37] Halvorsen RE, Elvestad M, Molin M, Aune D. Fruit and vegetable consumption and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis of prospective studies. BMJ Nutr Prev Health 2021;4:519−31.
[38] Aune D, Schlesinger S, Leitzmann MF, Tonstad S, Norat T, Riboli E, et al. Physical activity and the risk of heart failure: a systematic review and dose-response meta-analysis of prospective studies. Eur J Epidemiol 2021;36:367−81.
[39] Filippini T, Wise LA, Vinceti M. Cadmium exposure and risk of diabetes and prediabetes: a systematic review and dose-response meta-analysis. Environ Int 2022;158:106920.
[40] Yuan T, Zhang H, Chen B, Zhang H, Tao S. Association between lung cancer risk and inorganic arsenic concentration in drinking water: a dose-response meta-analysis. Toxicol Res 2018;7:1257−66.
[41] Fedirko V, Tramacere I, Bagnardi V, Rota M, Scotti L, Islami F, et al. Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. Ann Oncol 2011;22: 1958−72.
[42] Tsilas CS, de Souza RJ, Mejia SB, Mirrahimi A, Cozma AI, Jayalath VH, et al. Relation of total sugars, fructose and sucrose with incident type 2 diabetes: a systematic review and meta-analysis of prospective cohort studies. CMAJ 2017;189:E711−20.
[43] Morze J, Schwedhelm C, Bencic A, Hoffmann G, Boeing H, Przybylowicz K, et al. Chocolate and risk of chronic disease: a systematic review and dose-response meta-analysis. Eur J Nutr 2020;59: 389−97.
[44] Morgan RL, Thayer KA, Santesso N, Holloway AC, Blain R, Eftim SE, et al. A risk of bias instrument for non-randomized studies of exposures: a users' guide to its application in the context of GRADE. Environ Int 2019;122:168−84.

[45] Yu WW, Schmid CH, Lichtenstein AH, Lau J, Trikalinos TA. Empirical evaluation of meta-analytic approaches for nutrient and health outcome dose-response data. Res Synth Methods 2013;4:256−68.
[46] Potischman N, Weed DL. Causal criteria in nutritional epidemiology. Am J Clin Nutr 1999;69:1309S−14S.
[47] Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. J Clin Epidemiol 2017;87:4−13.
[48] Schunemann HJ. Interpreting GRADE's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? J Clin Epidemiol 2016;75:6−15.
[49] Zeng L, Brignardello-Petersen R, Hultcrantz M, Siemieniuk RAC, Santesso N, Traversy G, et al. GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings. J Clin Epidemiol 2021;137:163−75.
[50] Alper BS, Oettgen P, Kunnamo I, Iorio A, Ansari MT, Murad MH, et al. Defining certainty of net benefit: a GRADE concept paper. BMJ Open 2019;9:e027445.
[51] Ezzatvar Y, Ramirez-Velez R, Izquierdo M, Garcia-Hermoso A. Physical activity and risk of infection, severity and mortality of COVID-19: a systematic review and non-linear dose-response meta-analysis of data from 1 853 610 adults. Br J Sports Med 2022;56:1188−93.
[52] Morgan RL, Whaley P, Thayer KA, Schunemann HJ. Identifying the PECO: a framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. Environ Int 2018;121:1027−31.
[53] Mustafic H, Jabre P, Caussin C, Murad MH, Escolano S, Tafflet M, et al. Main air pollutants and myocardial infarction: a systematic review and meta-analysis. JAMA 2012;307:713−21.
[54] Verbeek JH, Whaley P, Morgan RL, Taylor KW, Rooney AA, Schwingshackl L, et al. An approach to quantifying the potential importance of residual confounding in systematic reviews of observational studies: a GRADE concept paper. Environ Int 2021;157:106868.
[55] Whaley P, Piggott T, Morgan RL, Hoffmann S, Tsaioun K, Schwingshackl L, et al. Biological plausibility in environmental health systematic reviews: a GRADE concept paper. J Clin Epidemiol 2022;146:32−46.