

SERIES

GRADE guidelines 27: how to calculate absolute effects for time-to-event outcomes in summary of findings tables and Evidence Profiles

Nicole Skoetz^{a,*}, Marius Goldkuhle^a, Elvira C. van Dalen^b, Elie A. Akl^c, Marialena Trivella^d, Reem A. Mustafa^e, Artur Nowak^f, Philipp Dahm^g, Holger Schünemann^h, Ralf Benderⁱ, GRADE Working Group

^aDepartment I of Internal Medicine, University of Cologne, Center for Integrated Oncology Aachen Bonn Cologne Duesseldorf, Kerpener Street 62, 50937 Cologne, Germany

^bPrincess Máxima Center for Pediatric Oncology, Heidelberglaan 25, 3584CS Utrecht, The Netherlands

^cDepartment of Internal Medicine, American University of Beirut, Lebanon

^dCentre for Statistics in Medicine, University of Oxford, Oxford, UK

^eDepartment of Medicine, University of Kansas Health System, 3901 Rainbow Blvd, MS3002, Kansas City, KS 66160, USA

^fEvidence Prime Inc, 175 Longwood Road South, Suite 305, Hamilton, Ontario L8P 0A1, Canada

^gMinneapolis VA Health Care System, Urology Section 112D, One Veterans Drive, Minneapolis, MN 55417, USA

^hDepartment of Health Research Methods, Evidence, and Impact, McMaster University, 1280 Main Street W, Hamilton, Ontario L8S 4K1, Canada

ⁱDepartment of Medical Biometry, Institute for Quality and Efficiency in Health Care, Im Mediapark 8, D-50670 Cologne, Germany

Accepted 10 October 2019; Published online 9 November 2019

Abstract

Objectives: To provide GRADE guidance on how to prepare Summary of Findings tables and Evidence Profiles for time-to-event outcomes with a focus on the calculation of the corresponding absolute effect estimates.

Study Design and Setting: This guidance was justified by a research project identifying frequent errors and limitations in the presentation of time-to-event outcomes in the Summary of Findings tables. We developed this guidance through an iterative process that included membership consultation, feedback, presentation, and discussion at meetings of the GRADE Working Group.

Results: Review authors need to carefully consider the definition of the outcome of interest; although often the event is used as label for the outcome of interest (e.g., death or mortality), the event-free survival (e.g., overall survival) is reported throughout individual studies. Review authors should calculate the absolute effect correctly, either for the event or absence of the event. We also provide examples on how to calculate the absolute effects for events and the absence of events for various baseline or control group risks and time points.

Conclusions: This article aids in the development of Summary of Findings tables and Evidence Profiles, including time-to-event outcomes, and addresses the most common scenarios when calculating absolute effects in order to provide an accurate interpretation. © 2019 Elsevier Inc. All rights reserved.

Keywords: GRADE guidance; Time-to-event outcomes; Hazard ratio; Absolute effects; Summary of findings table; Evidence Profile

1. Introduction

The GRADE approach provides a systematic and transparent framework for rating the certainty of evidence and moving from the evidence to a recommendation or decision. Therefore, the GRADE guidelines are highly relevant for systematic review authors, health technology assessment, and clinical practice guidelines developers [1]. The

assessment of the certainty of the evidence is presented in GRADE Summary of Findings tables or GRADE Evidence Profiles, together with absolute effect estimates for relative effects [2,3]. A recent methodological systematic review showed that review authors might have difficulties calculating absolute effects for time-to-event outcomes [4].

Analyses that assess the time to a given event for one or several groups of patients are used in clinical studies in some fields, in particular, oncology. These time-to-event analyses are valuable, particularly when the event of interest can occur at any point over an extended period of time and the time till event occurrence carries important value. A distinct feature of time-to-event analytic techniques is to

* Corresponding author. Tel.: +49221 478 96651; fax: +49221 478 96654.

E-mail address: nicole.skoetz@uk-koeln.de (N. Skoetz).

What is new?**Key findings**

- The GRADE Working Group describes the preferred approach for presenting absolute effects for time-to-event outcomes in Summary of Findings tables or Evidence Profiles and provides guidance on how to avoid common pitfalls.
- Systematic review authors should be cautious whether as event mortality (e.g., people being dead) or survival (e.g., people who are alive) was used in the considered publications.
- In most cases, the absolute effect will be estimated indirectly from the hazard ratio and an adequate baseline risk. If the estimation uncertainty of the baseline risk is a relevant source of the total estimation uncertainty, it should be taken into account in the estimation of the absolute effect.

What is the implication and what should change now?

- Systematic review authors and guideline developers are advised to use the herein presented approaches to derive and present absolute effects of time-to-event outcomes, in order to support clinical decision-making and healthcare recommendations whenever they use the GRADE approach.

incorporate the censored information, which refers to information from study participants who did not experience the event of interest during the follow-up period. To compare the effects of different interventions/management strategies on time-to-event outcomes between two groups, hazard ratios (HRs) with corresponding confidence intervals derived from Cox regression models are routinely calculated as the relative effect measure.

Although their use is not limited to the field of oncology, the resulting Kaplan Meier curves, also referred to as survival curves are closely associated with oncology. For patients with cancer, one of the most relevant outcomes is overall survival (OS). Progression-free survival (PFS), disease-free survival (DFS), and event-free survival are also often assessed outcomes as they provide complementary information to OS. In addition, time-to-event analyses can describe outcomes other than survival, such as time to hospital admission, time to passage of a ureteral stone, or time to the occurrence of specified adverse events. These examples are time-to-event outcomes, as they involve the assessment of both whether a particular event occurs, and also when it occurred [5].

Absolute effect estimates (i.e., risk difference, the number needed to treat) provide important supplementary

information to relative effect estimates by considering the control event rate over a given time period. As they take into account the underlying baseline risk for the event of interest in the study groups, absolute effect estimates are less vulnerable to exaggerated effect interpretation than relative effect estimates and allow a more appropriate assessment of the clinical relevance of effects [6]. Especially the absolute difference of events in both arms for one outcome at specific time points is essential for decision making and are a routine part of GRADE Summary of Findings tables and GRADE Evidence Profiles. They are automatically calculated by GRADE's official app GRADEpro GDT (grade.pro.org) [2,3]. The formula for calculation of the absolute effects varies depending on whether the relative effect estimate is a risk ratio or hazard ratio [7].

As mentioned above, a recent methodological systematic review showed that less than 30% of oncological Cochrane Reviews calculated absolute effects for time-to-event outcomes correctly and presented results in an easily interpretable way [4]. The main source of error is the confusion around whether the study authors describe the proportion of participants with a given event (e.g., death from any cause) or the proportion of participants who are event-free (e.g., overall survival). Furthermore, interpretation of results in the Summary of Findings tables was hampered by the lack of explanation of which baseline risk (BLR) review authors used to estimate the absolute effect or by entering incorrect numbers like the number of events instead of numbers of patients being event-free into the GRADEpro GDT software.

Given the above-described confusion and the lack of written GRADE guidance available on how to calculate absolute effects for time-to-event outcomes based on HR and how to avoid common pitfalls, the members of the GRADE time-to-event Working Group developed this guidance incorporating feedback from methodologists and stakeholders. The findings from a methodological review evaluating the presentation of absolute effects from time-to-event data in 97 cancer-related Cochrane reviews was presented first at the GRADE meeting in 2017, in Rome, Italy. This meeting was followed by two small group discussions during the GRADE biannual meetings in Cape Town, South Africa, 2017, and Bogota, Colombia, 2018, and one large group discussion in Manchester, UK, 2018, involving more than 80 international experts, where the paper was formally approved. To prepare the presentations and incorporate the feedback from the meetings, the group of authors met in a 60-minute online conference and came to a consensus regarding this GRADE guidance.

2. Direct calculation of the absolute effect

It should be noted that when individual participant data are available and if the risk difference is an appropriate measure of the meta-analysis, the absolute effect of an

intervention in individual trials can and should be estimated directly based upon the individual participant data and not indirectly via the estimates of the hazard ratio and the baseline risk. Therefore, in this case, not only is the estimation uncertainty of the hazard ratio taken into account, but the full uncertainty of the absolute effect estimate is also automatically taken into account.

In studies of time-to-event data, there is usually a staggered entry of patients into the study leading to varying follow-up times and censored observations. Sometimes, studies have a recruitment period with a staggered entry (say, for 1 year) and a fixed follow-up period for all patients, say for 2 years. In this case, you have complete observations for a period of 2 years. In the case of a single study with individual participant data and complete observation for all patients at least for a minimum time period, a specific time point with complete observations should be chosen, and the corresponding 2×2 table should be prepared. The usual methods for binary data can be applied to yield appropriate point and interval estimates for the risk difference [8].

In the case of staggered entry of the patients over the whole study duration, no adequate time period with complete observation may be available. In the case of a single study with individual participant data and incomplete observation, reviewers should apply methods for survival data. Methods based upon Kaplan-Meier curves [7,9] and the Cox regression model [10–12] are available to perform point and interval estimates for the risk difference at different time-points.

Sometimes it might be useful to choose the risk difference as effect measure for the meta-analysis (e.g., in the case of rare events and empty cells). In this case, a pooled risk difference by means of the usual meta-analytic methods represents an adequate measure of the absolute effect [13]. In all other cases, the estimation of the absolute effect should be performed indirectly from the pooled HR, and adequate estimation of the baseline risk.

3. Indirect calculation of the absolute effect

3.1. Assumptions for this guidance paper

For calculations of absolute effects from a pooled hazard ratio (HR), we assume that the latter is correctly calculated and applicable in the considered situation. Besides unadjusted HRs, the HRs adjusted for prognostic factors can also be used if the adjustment is performed adequately for the considered research question.

3.2. Assumptions to estimate baseline risks

The baseline risk used to calculate absolute effect size estimates should be appropriate for the individuals and their characteristics to which it is intended to be applied to. Typically, the calculation of absolute effects in systematic

reviews is based on the baseline risk from included trials. However, trials could enroll individuals with a higher than average baseline risk as a way to increase their statistical power, or they could include patients with a lower than average baseline risk, as patients with comorbidities might have been excluded.

3.3. Use of the baseline risk from an external source

Large, representative observational studies at low risk of bias or systematic reviews of those studies may provide adequate baseline event rates. This approach has been previously reported for binary outcomes using appropriate observational studies, with defined prognostic markers for different risk groups. If an appropriate estimate for the baseline risk with 95% confidence interval (CI) is available from an external source, for example, from an observational study or registry, it is possible to estimate the absolute effect by taking the uncertainties of the HR and the baseline risk estimates into account (see section 3.5). For representing multiple risk groups in the population, studies with different baseline risks could be grouped accordingly (i.e., into risk groups like high, moderate, and low). For each risk group, the baseline risk estimates of representative studies could then be used to calculate the corresponding absolute risks in the intervention arm. It must be noted that systematic review authors should not extrapolate the HR beyond the follow-up period that it represents. For example, if the (pooled) HR is calculated for a follow-up period of 1.5 years, a baseline risk at 1 year from an eligible observational study may be suitable to estimate a corresponding absolute risk at 1 year. Whereas, the same HR should not be extrapolated and applied together with a baseline risk for 5 years to estimate an absolute risk at 5 years. This is because we have evidence that HR is constant only within the period of 1.5 years. After this period, the HR could potentially increase or decrease.

3.4. Use of the baseline risk from the control groups of the included studies

If no suitable observational data are available to estimate the baseline risk, data from Kaplan-Meier survival curves from the control groups of the trials included in the systematic review may be used to estimate the baseline risk. An option here is to select the curve from a trial representative for the control group of interest that is estimated to be at low risk of bias. It is as well an option to choose the curves from multiple trials representing different baseline risk groups (e.g., high, moderate, and low). Again, as mentioned for observational studies, trials with different baseline risks could be grouped, and effect estimates of representative trials for each risk group could be used to calculate the absolute effect for the intervention arm.

Oftentimes, toward the end of the reported observation time, only a small number of patients may still be at risk,

with most patients having either experienced the event of interest or being censored. Therefore, the review authors should ideally choose a time point from the middle of the observation time of the respective Kaplan-Meier survival curve rather than at the end. This recommendation requires, however, that a sufficient number of events has happened up to the chosen time. In case there is a high degree of follow-up after this time point, meaning that none or only a few individuals are censored for a later eligible time point, it is possible to choose a later time point, where a larger number of events may have occurred. The chosen time point should be consistent across the different risk groups and clearly reported. Here again, it is important to point out that the HR should not be extrapolated and combined with a baseline risk estimate for a time-period that it does not represent (see section 3.3).

Sometimes, trials included in a meta-analysis, report on HRs only without presenting survival curves and survival rates at specific time points. In this case, no adequate estimates for the control group risk can be extracted from these trials and the observational data should be used to estimate the control group risk.

3.5. Uncertainty of the baseline risk estimate

Comparable to guidance for the calculation of absolute effects for binary data [14], only the uncertainty of the pooled HR is taken into account when grading the certainty of the body of evidence, not the uncertainty of the time point from the Kaplan-Meier survival curve and the corresponding baseline risk. The calculation of absolute effects is, therefore, conditional, based on the assumption that a given baseline risk is true.

As noted above, the baseline risk to estimate the absolute effect comes ideally from appropriate large, representative observational studies at low risk of bias. If this study is large, the standard error of the baseline risk estimate may be quite small, so that this uncertainty is negligible. In this case, the methods described in the next section can be used to estimate the absolute effect by using the baseline risk from the observational study.

However, in settings in which it appears important to take the uncertainty of the baseline risk estimate into account, which could be when the uncertainty of the baseline risk is a relevant source of the total uncertainty [15], a general method called Propagating Imprecision (PropImp) can be used to estimate the absolute effects [16]. Preconditions are that the baseline risk estimate comes from a source that is independent of the meta-analysis and that adequate point and confidence intervals are available for the baseline risk and the pooled HR. The computationally intensive PropImp approach is described in detail elsewhere, and an MS Excel sheet can be made available to facilitate implementation [16].

If large, representative observational studies at low risk of bias are not available, trials included in the meta-analysis may then provide the estimates of baseline risks

[14]. In this case, the uncertainties of the baseline risk and the relative effect are correlated. Thus, only complex methods, including resampling, are available to take the uncertainty of the baseline risk into account to construct a valid confidence interval for the absolute effect [17]. However, it is not always necessary to take the uncertainty of the baseline risk estimate into account. If the standard error of the baseline risk estimate is small and the standard error of the pooled HR is the main source of the total uncertainty, the uncertainty of the baseline risk estimate is negligible. Under these circumstances, we can also take up the conditional view. Especially if we calculate the absolute effect for different risk groups, it makes sense to present the various absolute effects conditional on the corresponding assumed baseline risks. In this case, it is sufficient to take only the uncertainty of the pooled HR into account.

3.6. Transparent reporting

As suggested by Santesso et al. [18], transparent reporting of where baseline risk data come from is very important. It should be clearly described in the explanatory footnotes where the baseline risk comes from and which specific time point has been chosen. The time-to-event outcome and the corresponding absolute effects in the Summary of Findings table or Evidence Profile should be labeled in a consistent manner throughout the review (e.g., in the abstract, methods, and results section). The reviewers need to make a clear distinction between people who are event-free (e.g., people alive at a specific time point) and people with an event (e.g., people dead at a specific time point). If both, events and absence of events are reported in different sections of the review, a clear explanation is needed to avoid confusing the reader.

The calculated absolute effects should be reported in the Summary of Findings table and in addition at least in the abstract [19], as absolute effect estimates are more understandable to patients, clinicians, and other users of evidence syntheses than relative effect measures and are the recommended effect measure to communicate risks [20].

The specific time point of the baseline risk, which was used to calculate an absolute effect size estimate, should be provided rather than time ranges. Sometimes, review authors use the total number of events observed across several included trials with different follow-up durations to inform the baseline risk to estimate the corresponding absolute effect size estimate. This is not helpful to users since clinical decision-making is based on effect size estimates at a certain time point (e.g., 5 years or 60 months), and absolute effect size estimates will vary greatly depending on the time-point chosen.

4. Estimating and presenting absolute effects

First, we suggest to clearly define what is meant by event (e.g., people being dead) or by event-free survival (e.g.,

Hazard Ratio (HR) is a time event measure of relative effect, estimated in survival analysis. It is calculated for an event (e.g. death) but the absolute effect (e.g. risk difference) has been customarily presented as either reduction/increase of a risk of an event (e.g. Death) or as an improvement/deterioration of non event (e.g. survival).

GRADE HANDBOOK provides [more info](#)

Which category best describes this outcome: "New outcome"

- An event (e.g. death, exacerbation)
- An non-event (commonly event-free survival)

Cancel

Save

Fig. 1. Options to determine the definition and category of the event of interest (event [cumulative incidence] or non-event [survival]) of a time-to-event outcome in the GRADEpro GDT software, which is used to create Summary of Findings tables and Evidence Profiles.

people who are alive) and to estimate the desired proportion by labeling clearly whether this is the proportion of patients with event or patients being event-free (please see Fig. 1).

4.1. Calculations of absolute effects for event-free survival (e.g., overall survival, progression-free survival)

Calculation of absolute effects is based on methods as described by Tierney et al. [5] under the assumption of proportional hazards. Let p_i , $i = 0,1$, be the proportion of event-free patients up to a given time point in the control ($i = 0$) and intervention group ($i = 1$), respectively, and HR the hazard ratio for the comparison of the hazard between the intervention and the control group (intervention/control). Then the proportion of event-free patients in the intervention group can be calculated as:

$$p_1 = \exp(\ln(p_0) \times HR) = p_0^{HR}$$

As an example, a pooled HR of 0.42 (95% CI 0.25 to 0.72) is used, indicating a lower risk of death over time in the intervention group. Estimating a proportion of

patients with event-free survival in the control group at the time point 2 years of 0.9 we obtain:

$$p_1 = \exp(\ln(0.9) \times 0.42) = .9^{0.42} = 0.957.$$

This means that 96 of 100 people with this disease will be alive with the experimental intervention at 2 years. Then, the upper and lower confidence limits for the corresponding intervention risk are obtained by replacing HR by their upper and lower confidence limits, respectively (e.g., replacing 0.42 with 0.25, then with 0.72, in the example above), according to the substitution method of Daly (please see Fig. 2) [22].

4.2. Calculation of absolute effects for events (e.g., mortality)

For obtaining absolute effects for time-to-event outcomes reported as events, such as mortality, a similar formula can be used. Let r_i , $i = 0,1$, be the proportion of patients with event up to a given time point in the control ($i = 0$) and intervention group ($i = 1$), respectively (i.e.,

Outcome	Anticipated absolute effects (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty
	Risk with no preoperative chemotherapy	Risk with preoperative chemotherapy			
Overall survival follow up: 2 years	Low		HR 0.87 (0.78 to 0.96) [survival]	2385 (15 RCTs)	⊕⊕⊕⊕ HIGH
	55 per 100	59 per 100 (56 to 63)			
Overall survival follow up: 5 years	Low		HR 0.87 (0.78 to 0.96) [Survival]	2385 (15 RCTs)	⊕⊕⊕⊕ HIGH
	40 per 100	45 per 100 (41 to 49)			

Fig. 2. Example: Calculations for event-free survival (overall survival) at two time points, based on an example in lung cancer patients [21]. In this example, an HR < 1 favors the intervention group, so more people will be alive in the intervention arm compared to the control arm. Please note that the term “risk” in the column headings misleadingly addresses the “risk” of surviving.

Outcome	Anticipated absolute effects (95% CI)		Relative effect (95% CI)	No of participants (studies)	Certainty
	Risk with no preoperative chemotherapy	Risk with preoperative chemotherapy			
Mortality follow up: 2 years	Low		HR 0.87 (0.78 to 0.96) [Mortality]	2385 (15 RCTs)	⊕⊕⊕⊕ HIGH
	45 per 100	41 per 100 (37 to 44)			
Mortality follow up: 5 years	Low		HR 0.87 (0.78 to 0.96) [Mortality]	2385 (15 RCTs)	⊕⊕⊕⊕ HIGH
	60 per 100	55 per 100 (51 to 59)			

Fig. 3. Example: Calculations for events (mortality) at two time points, based on an example in lung cancer patients [21]. In this example, an HR < 1 favors the intervention group, so fewer people will be dead in the intervention arm compared to the control arm.

r_0 is the baseline risk), then risk of an event in the intervention group can be calculated by

$$r_1 = 1 - \exp(\ln(1 - r_0) \times HR) = 1 - (1 - r_0)^{HR}$$

Fig. 3 gives an example of the presentation in GRADEpro.

4.3. Graphical presentation

For supporting the interpretation of systematic review results, the GRADEpro software provides the opportunity to present review findings graphically in an interactive summary of findings table [23]. A feature of this format allows visualizing a corresponding absolute effect for the comparison of an intervention arm to a control arm for each outcome. In six steps, the absolute number of events for a specific time point in the control group (the baseline risk), the estimated number of events in the intervention group, the risk difference and the associated statistical uncertainty are presented in an easily comprehensible way (please see Fig. 4 for an example).

4.4. Calculation of numbers needed to treat based on events or event-free survival

Numbers needed to treat with confidence intervals can also be calculated as the inverse of the risk differences between intervention and control arm [24].

Risk difference: control group risk–intervention group risk (95% CI [control group risk–upper CI; control group risk–lower CI])

Example from above for events (mortality at 2 years)

Risk difference: 45/100 (control group)–41/100 (95% CI 37 to 44) (intervention group) = 4/100 (95% CI 1/100 to 8/100)

1/Risk difference = 25 (95% CI 12.5 to 100)

Meaning that 25 (13 to 100) people need to be treated to avoid one death at 2 years.

Similar to the afore outlined calculations of absolute effects utilizing the HR and corresponding baseline risk, the number needed to treat is strongly depending on the size of the chosen baseline risk [25]. Therefore, here, we propose to present the numbers needed to treat and the corresponding upper and lower confidence intervals

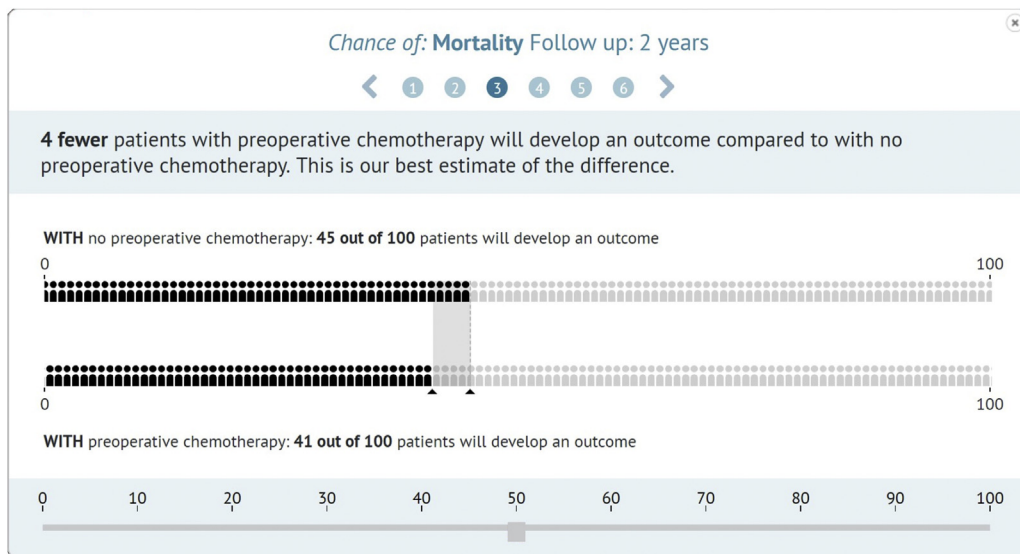


Fig. 4. Example: Graphical presentation of the absolute number of events in the control and the intervention arm at 2 yr in the interactive Summary of Findings table, based on an example in lung cancer patients [21].

across a range of baseline risks to represent different risk groups.

4.5. Calculation of median survival time

Calculation of median (event-free) survival time while applying the HR is one of the options presented in the paper by Tierney et al. [5] for individual trials. This option might be of great interest to patients, physicians, and stakeholders for clinical decision-making, but user testing is needed. One necessary condition is that the median survival time has been reached in the control group, meaning that for overall survival, 50% of the patients at risk already died. For obtaining the median survival time in the intervention group (MST_1) from the median survival time in the control group (MST_0) and the pooled HR, the following formula can be used (the calculation is based upon the assumption that MST_0 is fixed):

$$MST_1 = MST_0 / HR$$

As an example, we consider the pooled hazard ratio of $HR = 0.42$ (95% CI 0.25 to 0.72). In this case, $HR < 1$ is defined as favoring the intervention arm. Assuming a median survival time in the control group of 80 months, we obtain:

$$MST_1 = 80 \text{ months} / 0.42 = 190.5 \text{ months}$$

Again, only the uncertainty of the HR is taken into account, not that of the median survival time. Upper and lower confidence limits for the corresponding intervention risk are obtained by replacing HR by their upper and lower confidence limits, respectively (e.g., replacing 0.42 with 0.25, then with 0.72, in the example above).

The difference of the median survival times between the intervention and the control group can be calculated by $MST_1 - MST_0 = 190.5 \text{ months} - 80 \text{ months} = 110.5 \text{ months}$.

5. Summary

Absolute effect estimates, especially absolute risk differences, provide essential information to guide clinical decision-making and the formulation of healthcare recommendations.

For time-to-event outcomes, the GRADE approach focuses on absolute effect estimates that are calculable from a hazard ratio and an applicable baseline risk as these will most frequently be available to the systematic review and guideline authors. Thus, GRADE focusses on risk differences and, on occasion, the number needed to treat or median survival times. We here present several approaches that are suitable to calculate the corresponding estimates and accord to the available data. In situations where sufficient data (e.g., IPD) or complete information for all study participants for a fixed follow-up duration is available, we advise review authors to use direct estimation methods, which are outlined in this document. As these data are often

not available, we also guide review or guideline authors on how to calculate absolute effects indirectly.

When calculating absolute effect estimates, review authors must consider the direction of the effect (which intervention is favored with an $HR < 1$?) and whether the cumulative incidence of the event or event-free survival is reported, as given by the definition of the outcome. Authors, as well as users of systematic reviews, should be aware of potential mistakes in the calculation of absolute effects and should include the direction of the relative effect into their judgment.

6. Further considerations and unresolved issues

The GRADEpro GDT software has been adapted to provide systematic review authors and guideline developers the opportunity to choose from the number of people with a given event or without an event at a specific time point when presenting absolute effect size estimates. This allows consistency of reported outcomes throughout the review and lets authors and guideline developers choose the format that seems most suitable to questions at hand.

In this guidance paper, we focused only on the correct calculation of absolute effects and interpretation of the direction of effect—event vs. event-free survival. There are a number of unresolved issues related to meta-analyses of time-to-event outcomes and grading the certainty of the evidence body. The GRADE Working Group is aiming to address the following issues in subsequent guidance:

- Time-to-event outcomes have features that typically incorporate observations based on censoring [26]. Further challenging aspects are to assess the certainty of the evidence for censoring mechanisms that are not independent of the outcome leading to a potential risk of bias.
- Treatment-switching is nowadays common in cancer trials, which also might introduce bias in time-to-event analyses. Assessment of this bias is particularly difficult as the time points of switching are usually not given. How to grade the certainty of the evidence in case of treatment switching will be elucidated in another guidance paper
- In competing risk settings, sometimes Kaplan-Meier survival analyses are performed, which might overestimate a potential effect [27], which will also be the focus of another paper.
- In cases where the proportional hazards assumption is invalid, alternative effect measures to the HR, such as the difference of the restricted mean survival time (RSMT) between the groups, can be used [28].

CRedit authorship contribution statement

Nicole Skoetz: Writing - original draft, Methodology, Writing - review & editing, Project administration,

Conceptualization, Supervision. **Marius Goldkuhle:** Writing - original draft, Writing - review & editing, Methodology, Project administration, Conceptualization. **Elvira C. van Dalen:** Writing - original draft, Writing - review & editing, Methodology, Conceptualization. **Elie A. Akl:** Writing - original draft, Writing - review & editing, Methodology. **Marialena Trivella:** Writing - original draft, Writing - review & editing, Methodology. **Reem A. Mustafa:** Writing - original draft, Writing - review & editing, Methodology. **Artur Nowak:** Writing - original draft, Writing - review & editing, Software. **Philipp Dahm:** Writing - original draft, Writing - review & editing, Methodology. **Holger Schünemann:** Writing - original draft, Methodology, Writing - review & editing. **Ralf Bender:** Writing - original draft, Writing - review & editing, Methodology, Project administration, Conceptualization, Supervision.

Acknowledgments

Article history: Slides presented at GRADE meetings in Rome (2017), Cape Town (2017), Bogota (2018) and Manchester (2018).

References

- [1] Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *J Clin Epidemiol* 2011;64:380–2.
- [2] Schünemann HJ, Vist GE, Glasziou P, Akl E, Skoetz N, Guyatt GH. Chapter 14: completing summary of findings tables and grading the certainty of evidence. In: Higgins JPT, Chandler J, Cumpston M, Li T, Page MJ, Welch V, editors. *Cochrane handbook for systematic reviews of interventions version 6* (updated January 29, 2019). Chichester (UK): The Cochrane Collaboration; 2019:2019. Available at <https://training.cochrane.org/handbooks>. Accessed September 19, 2019.
- [3] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383–94.
- [4] Skoetz N, Goldkuhle M, Weigl A, Dwan K, Lebono V, Dahm P, et al. Methodological review showed correct absolute effect size estimates for time-to-event outcomes in less than one-third of cancer-related systematic reviews. *J Clin Epidemiol* 2019;108:1–9.
- [5] Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007;8:16.
- [6] Carling CLL, Kristoffersen DT, Montori VM, Herrin J, Schünemann HJ, Treweek S, et al. The effect of alternative summary statistics for communicating risk reduction on decisions about taking statins: a randomized trial. *PLoS Med* 2009;6(8):e1000134.
- [7] Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999;319(7223):1492–5.
- [8] Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med* 1998;17:873–90.
- [9] Bender R, Kromp M, Kiefer C, Sturtz S. Absolute risks rather than incidence rates should be used to estimate the number needed to treat from time-to-event data. *J Clin Epidemiol* 2013;66:1038–44.
- [10] Austin PC. Absolute risk reductions and numbers needed to treat can be obtained from adjusted survival models for time-to-event outcomes. *J Clin Epidemiol* 2010;63:46–55.
- [11] Laubender RP, Bender R. A note on calculating asymptotic confidence intervals for the adjusted risk difference and number needed to treat in the Cox regression model. *Stat Med* 2014;33:798–810.
- [12] Laubender RP, Bender R. Estimating adjusted risk difference (RD) and number needed to treat (NNT) measures in the Cox regression model. *Stat Med* 2010;29:851–9.
- [13] Deeks J, Higgins JPT, Altman D, on behalf of the Cochrane Statistical Methods Group. Chapter 9: analysing data and undertaking meta-analyses. In: Higgins JPT, Churchill R, Chandler J, Cumpston MS, editors. *Cochrane handbook for systematic reviews of interventions version 5.2.0* (updated June 2017). Cochrane: Cochrane; 2017:2017.
- [14] Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, et al. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. *J Clin Epidemiol* 2013;66:158–72.
- [15] Spencer FA, Iorio A, You J, Murad MH, Schunemann HJ, Vandvik PO, et al. Uncertainties in baseline risk estimates and confidence in treatment effects. *BMJ* 2012;345:e7401.
- [16] Newcombe RG. Propagating imprecision: combining confidence intervals from independent sources. *Commun Stat - Theor Methods* 2011;40(17):3154–80.
- [17] Newcombe RG, Bender R. Implementing GRADE: calculating the risk difference from the baseline risk and the relative risk. *Evid Based Med* 2014;19(1):6–8.
- [18] Santesso N, Carrasco-Labra A, Langendam M, Brignardello-Petersen R, Mustafa RA, Heus P, et al. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. *J Clin Epidemiol* 2016;74:28–39.
- [19] Agarwal A, Johnston BC, Vernooij RW, Carrasco-Labra A, Brignardello-Petersen R, Neumann I, et al. Authors seldom report the most patient-important outcomes and absolute effect measures in systematic review abstracts. *J Clin Epidemiol* 2017;81:3–12.
- [20] Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping Doctors and patients make sense of health statistics. *Psychol Sci Public Interest* 2007;8(2):53–96.
- [21] Preoperative chemotherapy for non-small-cell lung cancer: a systematic review and meta-analysis of individual participant data. *Lancet* 2014;383(9928):1561–71.
- [22] Daly LE. Confidence limits made easy: interval estimation using a substitution method. *Am J Epidemiol* 1998;147:783–90.
- [23] Schunemann HJ, Mustafa R, Brozek J, Santesso N, Alonso-Coello P, Guyatt G, et al. GRADE Guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health. *J Clin Epidemiol* 2016;76:89–98.
- [24] Veroniki AA, Bender R, Glasziou P, Straus SE, Tricco AC. The number needed to treat in pairwise and network meta-analysis and its graphical representation. *J Clin Epidemiol* 2019;111:11–22.
- [25] Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses—sometimes informative, usually misleading. *BMJ* 1999;318(7197):1548–51.
- [26] Kleinbaum DG, Klein M. *Survival analysis. A self-learning text*. New York: Springer; 2012. Available at <http://www.springer.com/de/book/9781441966452>. Accessed September 19, 2019.
- [27] Lacny S, Wilson T, Clement F, Roberts DJ, Faris P, Ghali WA, et al. Kaplan-Meier survival analysis overestimates cumulative incidence of health-related events in competing risk settings: a meta-analysis. *J Clin Epidemiol* 2018;93:25–35.
- [28] Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol* 2013;13:152.