GRADEprofiler help

1. Introduction to GRADE Handbook

GRADE handbook

GRADE handbook

for grading the quality of evidence and the strength of recommendations

Version 3.2

[updated March 2009]

Editors: Holger Schünemann, Jan Brożek, Andrew Oxman

Handbook Information

1.1 Handbook Information

Handbook Information

About the Handbook

The GRADE handbook describes the use of the GRADEprofiler (GRADEpro), a tool used to apply the GRADE approach to grading the quality of evidence and strength of recommendations.

Keeping up to date

The Handbook is updated regularly to reflect advances in the GRADE approach based on response to feedback from users and members of the GRADE Working Group. It is based on the third, and completely revised version of GRADEpro. Please refer to http://www.cc-ims.net/gradepro for the most recent version, for updates to the guidance. Users of the Handbook are encouraged to send feedback and corrections to the Handbook editors; please refer to the web site.

Reproduction and translation

Permission to reproduce or translate the GRADE handbook for grading the quality of evidence and the strength of recommendation should be sought from the editors; please refer to the web site.

How to cite this version of the handbook

Schünemann H, Brozek J, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendation. Version 3.2 [updated March 2009]. The GRADE Working Group, 2009. Available from http://www.cc-ims.net/gradepro.

When referring to a specific section or subsection refer to it by the title and section number, not page numbers. Example:

Schünemann H, Brozek J, Oxman A, editors. GRADE handbook for Grading quality of evidence and strength of recommendation. 3.2 [updated March 2009]. Section 3

Sources of Support

- **European Commission**
  The human factor, mobility and Marie Curie Actions Reintegration Grants: IGR 42192 — "GRADE"
- **Norwegian Knowledge Centre for the Health Services**

Internet address

www.gradeworkinggroup.org

Acknowledgements

We are grateful to all past contributors including those who have provided feedback on the early versions of the handbook, and would particularly like to thank **Roman Jaeschke, Robin Habour** and **Elie Akl** (major revision to version 2.0).

Major Contributors

Handbook Editors

Holger Schünemann, University at Buffalo, New York, USA & McMaster University, Hamilton, Canada
Jan Brozek, Polish Center for Evidence Based Medicine, Kraków, Poland
Andrew Oxman, Norwegian Knowledge Centre for the Health Services, Oslo, Norway

Authors and Editors

The members of the GRADE Working Group have contributed to this handbook.
The following authors have made major contributions: **Gordon Guyatt, Nancy Santesso, Elie Akl**

v.090311.2024

2. Welcome to the GRADEprofiler

Welcome to the GRADEprofiler (GRADEpro)

GRADEprofiler (GRADEpro) is an application for creating, managing, and sharing evidence profiles and Summary of Findings Tables.

This tool is intended for:

- authors of systematic reviews
- guideline developers
- those requiring summaries of the best available evidence for recommendations about particular courses of action in health care

---

Table of contents

- Managing evidence profiles
- Exporting evidence tables

For most features GRADEpro provides a context-specific help. Please use it whenever in doubt.

3. What is new?

What is new in this version of GRADEprofiler?

In future releases in this section you will find a **summary of most important changes**.

Please visit this section each time you download a new version of GRADEprofiler.

4. Getting started

Learning Basics About the GRADEprofiler

- About evidence tables
- Tasks you can accomplish
- User interface
- Context-specific help
- Keyboard shortcuts

4.1 Tasks you can accomplish

Tasks One Can Accomplish with GRADEprofiler

GRADEpro allows the creation and export of evidence profiles in various ways.

Overview of tasks that can be accomplished with GRADEpro

- GRADE evidence profile
- Cochrane Collaboration Summary of Findings table
- Cochrane Collaboration Overview of Reviews table

Regardless of which table tyou create, you can either

- import data for Review Manager 5.0 file and edit them
- enter all data manually

4.2 Evidence tables

About Evidence Tables

GRADEpro lets you create three types (layouts) of evidence tables (summaries).

Three types of evidence tables are available

- GRADE evidence profile
- Cochrane Collaboration Summary of Findings (SoF) table
- Cochrane Collaboration Overview of Reviews table

Each type of evidence profile presents information in different way.

4.2.1 Evidence tables - general

About Evidence Tables

An **evidence table** is a key tool in the presentation of evidence and the corresponding results. It displays the information about all outcomes a for a given health care question in tabular format. This information should usually come from systematic reviews.

Clinicians, patients, guideline developers, and policy-makers require such succinct and transparent evidence summaries to support their decisions. While an unambiguous health care question is key to evidence summaries, the requirements for specific users may differ in content and detail.

Three layouts of evidence tables are available:

- GRADE evidence profile
- Summary of Findings (SoF) table
- Overview of reviews table

GRADE evidence profiles contain detailed information about the quality of evidence and the summary of findings for each of the included outcomes.

Summary of Findings (SoF) tables provide key pieces of information regarding the results of the available evidence and a summary of the quality of evidence for each outcome in a quick and accessible format. Authors of Cochrane systematic reviews are strongly encouraged to include SoF tables in their reviews, and to ensure that there is sufficient description of the studies and meta-analyses to support their contents.

Overview of Reviews tables provide key information about available evidence for related questions in a summary format (*e.g.* all interventions for a given health problem).

4.2.2 GRADE evidence profile

About GRADE Evidence Profile

A GRADE evidence profile allows presentation of key information about all relevant outcomes for a given health care question. It presents **information about the body of evidence** (*e.g.* number of studies), the **judgments about the underlying quality of evidence**, key **statistical results**, and a **grade for the quality of evidence for each outcome**.

A GRADE evidence profile is **particularly useful** for presentation of evidence supporting a recommendation in **clinical practice guidelines** but also as summary of evidence for other purposes where users need or want to understand the **judgments about the quality of evidence** in more detail. After completing the information for an evidence profile or another evidence table, the information will be stored and can be updated accordingly.

Example

Date: 2007-09-05
Question: Should compression stockings vs without stockings be used in anyone taking a long flight (lasting more than 6 hours)?
Settings: International air travel
Bibliography: Clarke MJ, Hopewell S, Juszczak E, Eisinga A, Kjeldstrøm M. Compression stockings for preventing deep vein thrombosis in airline passengers. Cochrane Database of Systematic Reviews [Year], Issue [Issue].

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	compression stockings[4]	without stockings	Relative (95% CI)	Absolute		
Symptomatic deep vein thrombosis (DVT)												
9	randomised trial					none	0/2821	0/0	not pooled	not pooled		CRITICAL
Symptom-less deep vein thrombosis (DVT)												
9	randomised trial	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	none	3/1314	47/1323 3%	OR 0.1 (0.04 to 0.25)	8 fewer per 1,000 26 fewer per 1,000	⊕⊕⊕⊕ HIGH	IMPORTANT
Superficial vein thrombosis												
8	randomised trial	no serious limitations	no serious inconsistency	no serious indirectness	serious[1]	none	4/903	12/901	OR 0.45 (0.18 to 1.13)	7 fewer per 1000 (from 11 fewer to 2 more)	⊕⊕⊕O MODERATE	IMPORTANT
Oedema (measured with: Post-flight values (oedema tests, ankle circumference and volume, and swelling and discomfort as assessed by the participant)[2]; range of scores: 0-10; Better indicated by less)												
6	randomised trial	very serious[3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	625	621	-	MD -4.7 (-4.9 to -4.5)	⊕⊕OO LOW	IMPORTANT

4.2.3 Summary of Findings table

About Summary of Findings Table

Summary of Findings (SoF) tables present the main findings of a systematic review in a transparent and simple tabular format. A Summary of Findings table provides key information concerning the quality of evidence, the magnitude of effect of the interventions examined, and the sum of available data on most important outcomes.

In contrast to a GRADE evidence profile, SoF tables present only the final grade for the quality of evidence for each outcome. Nevertheless, users of GRADEpro will have to make these judgments prior to obtaining a grade for the quality of evidence.

General template for Summary of Findings tables

The format of SoF tables produced using GRADEpro has been developed over the past several years through wide consultation, user testing, and evaluation. It is designed to support the optimal presentation of the key findings of systematic reviews. This format for SoF table has been developed with the aim of ensuring consistency and ease of use across reviews, inclusion of the most

important information needed by decision makers, and optimal presentation of this information. However, there may be good reasons for modifying the format of a SoF table for some reviews.

Standard SoF tables include seven columns:

- **outcomes** — list of important desirable and undesirable outcomes (SoF restricts the number of outcomes to seven most important ones)
- **assumed risk** — a measure of the typical burden of these outcomes, *i.e.* illustrative risk or illustrative mean of control intervention also called baseline risk, baseline score (final values or change score in control group), or control event rate
- **corresponding risk** — a measure of the burden of the outcomes after the intervention is applied, *i.e.* the risk of an outcome in treated/exposed people based on the relative magnitude of an effect and assumed (baseline) risk
- **relative magnitude of effect** — for dichotomous outcomes the table will usually provide risk ratio, odds ratio, or hazard ratio
- **number of participants and studies** addressing these outcomes
- **rating of the overall quality of evidence** for each outcome (which may vary by outcome)
- **comments** (if needed)

Where possible, **it is desirable to provide both relative and absolute measures of effect**. For types of data other than dichotomous, SoF tables will display either an absolute measure alone (such as difference in means for continuous data) or a relative measure alone (*e.g.* hazard ratio for time-to-event data).

Reviews with more than one main comparison require **separate SoF tables** for each comparison.

Example

Compression stockings compared with no compression stockings for people taking long flights

Patient or population: anyone taking a long flight (lasting more than 6 hours)
Settings: International air travel
Intervention: compression stockings[1]
Comparison: without stockings

Outcomes	Illustrative comparative risks* (95% CI) Assumed risk without stockings	Corresponding risk compression stockings	Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
Symptomatic deep vein thrombosis (DVT)	See comment	See comment	Not estimable	-	See comment	0 participants developed symptomatic DVT in these studies.
Symptom-less deep vein thrombosis (DVT)	**Low risk population[2]**		**OR 0.1** (0.04 to 0.25)	2637 (9)	⊕⊕⊕⊕ **high**	
	10 per 1000	**1 per 1000** (0 to 3)				
	High risk population[2]					
	30 per 1000	**3 per 1000** (1 to 8)				
Superficial vein thrombosis	**13 per 1000**	**6 per 1000** (2 to 15)	**OR 0.45** (0.18 to 1.13)	1804 (8)	⊕⊕⊕O **moderate[3]**	
Oedema Post-flight values (oedema tests, ankle circumference and volume, and swelling and discomfort as assessed by the participant)[4]. Scale from: 0 to 10.	The mean oedema ranged across control groups from **6 to 9 points**	The mean Oedema in the intervention groups was **4.7 lower** (4.9 to 4.5 lower)		1246 (6)	⊕⊕OO **low[5]**	

4.2.4 Overview of Reviews table

About Overview of Reviews Table

The **Overview of Reviews** table presents the key questions and outcomes of an Overview of Reviews (*e.g.* umbrella reviews) designed to compile evidence from multiple systematic reviews of interventions into one accessible and usable document. The column headings of these tables correspond to the outcomes assessed in the reviews and the related information regarding the number of participants and studies that were included in each of the comparisons for a specific outcome. As for the SoF table, up to seven outcomes should be included. The rows describe the comparisons addressed in the reviews.

Example

Outcome	Symptomatic deep vein thrombosis (DVT)	Symptom-less deep vein thrombosis (DVT)	Superficial vein thrombosis	Oedema	Pulmonary embolus	Death	Adverse effects
Comparisons	Number of participants (Number of studies)						
compression stockings vs without stockings in anyone taking a long flight (lasting more than 6 hours)	2821 (9)	2637 (9)	1804 (8)	1246 (6)	2821 (9)	2821 (9)	1182 (4)

Tip for authors of systematic reviews

Overview of Reviews (*e.g.* umbrella reviews) are intended primarily to overview multiple Cochrane Intervention reviews addressing the effects of two or more potential interventions for a single condition or health problem. Cochrane Overviews highlight the Cochrane reviews that address these potential interventions and summarize their results for important outcomes. One aim of an Overview of Reviews is to allow the Cochrane Library reader a quick overview of Cochrane reviews relevant to the clinical decision at hand (Chapter 22 of the Cochrane Handbook deals with Cochrane Overviews).

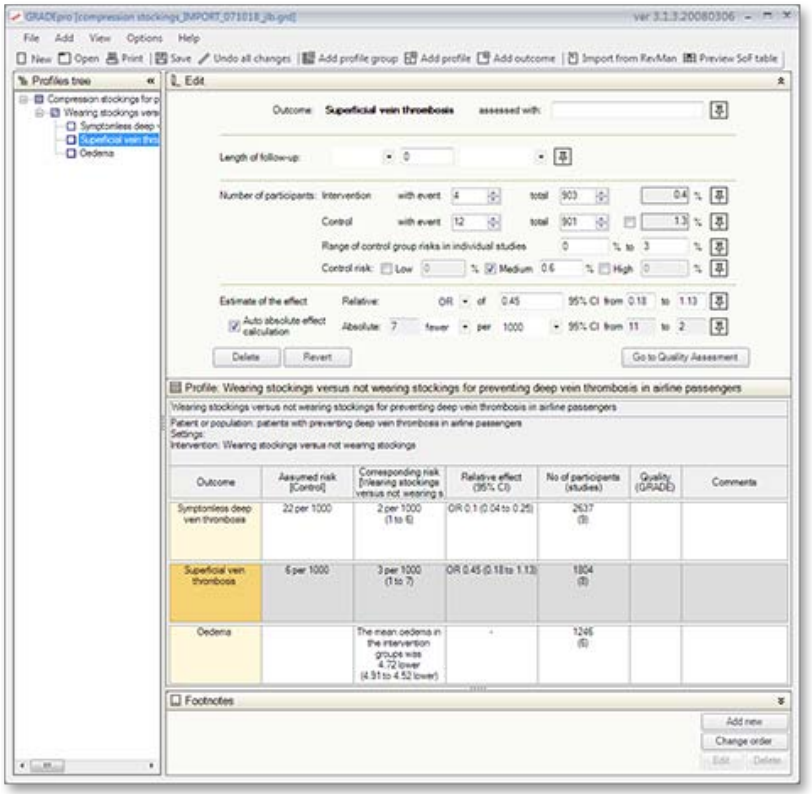
Authors should mention if an outcome has not been addressed for a specific question (indicated by 0 studies).

4.3 User interface

About GRADEprofiler User Interface

The GRADEpro user interface provides easy access to the various tools and features you can use to create, view, edit, manage, and export your evidence profiles and (SoF) Tables. The main areas of the GRADEpro interface are:

1. Menu commands
2. Toolbar
3. Tree pane of outcomes grouped into profiles (may be minimized)
4. Input and edit pane (may be minimized)
5. Preview of evidence table pane
6. Footnotes pane (may be minimized)



4.3.1 Menu commands

About menu Commands

- File
- Add
- View
- Options
- Help

Copyright © 2004–2009 GRADE Working Group

4.3.1.1 File

File menu

MENU ITEM	ACTION
New	Creates a new GRADEpro document. It will appear in a new GRADEpro window.
Open...	Opens an existing GRADEpro document. It will appear in a new GRADEpro window.
Open recent ...	Opens a submenu in which recently opened GRADEpro documents are listed. You may choose one of the recently opened documents from the list. It will appear in a new GRADEpro window.
Save	Saves changes to current document. GRADEprofiler saves your document everytime you press save button in the edit pane and asks you to save changes when you quit.
Save as...	Saves a copy of a currently open document.
Export as...	Exports selected group(s) of profiles, profile(s), or outcome(s) as: ■ Review Manager Summary of Findings (SoF) table (may be imported into RevMan 5.0) ■ Microsoft® Office Word document (may be further edited and used in a publication) ■ HTML document (may be published online or imported to a word processor) ■ Image (available bitmap image formats include: JPEG, GIF, BMP, and PNG) ■ XML file
Import from...	Imports selected group(s) of profiles, profile(s), or outcome(s) from: ■ Review Manager 5.0 file (*.rm5) ■ GRADEprofiler file (*.grd) ■ files generated with previous versions of GRADEpro [1.x and 2.x] (*.xml)
Print	Prints selected profile(s) or outcome(s) to a printer.
Quit	Exits GRADEpro.

4.3.1.2 Add

Add menu

MENU ITEM	ACTION
Profile Group	Creates a new Profile Group.
Profile	Creates a new Profile.
Outcome	Creates a new Outcome.

4.3.1.3 View

View menu

MENU ITEM	ACTION
Single row	Changes current preview to GRADE evidence table displayed in single row.
Double row	Changes current preview to GRADE evidence table displayed in two rows: upper one being quality of evidence and lower one being summary of findings.
Summary of Findings	Changes current preview to Cochrane Summary of Findings table.

4.3.1.4 Options

Options menu

MENU ITEM	ACTION
Force footnotes	Toggles forced footnotes on or off.
Autosave	Toggles autosaving on or off. When autosave option is selected all changes to the GRADEpro file you make are automatically saved every 5 minutes. When autosave option is deselected no changes are saved until you press **SAVE** button or **Ctrl-S**.
Show startup panel	Toggles startup panel (welcome screen) on or off.
Styles	Lets you change the visual appearance of the interface. This option may be helpful if your visual style of the Windows interface "skins" GRADEprofiler's user interface making it less ergonomic (e.g. low contrast of selected fields etc.).

4.3.1.5 Help

Help menu

MENU ITEM	ACTION
Contents	Lets you browse help table of contents.
Context specific help	Opens this help item explaining the use of context specific help.
Index	Lets you browse and search help index.

Search	Lets you search help file by keywords.
Website	Takes you to GARDE Working Group's website (www.gradeworkinggroup.org)
About	Displays GRADEprofiler copyright statement, version number, and credits.

4.3.2 Toolbar icons

Toolbar icons

Toolbar lets you quickly access most frequently used commands.

ICON	DESCRIPTION
	New GRADEpro file
	Open an existing GRADEpro file
	Print...
	Save current file
	Undo
	Add new Group of Profiles
	Add new Profile
	Add new Outcome
	Import from RevMan
	Preview SoF table

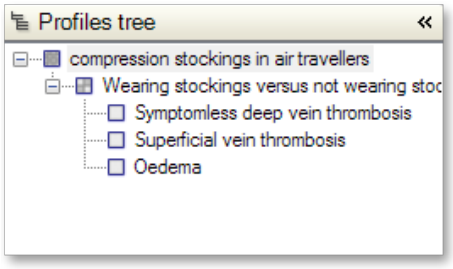
4.3.3 Tree pane

About the Tree of Profiles

The **tree of profiles** is displayed in the **tree pane** on the left. It showes the data structure of each GRADEpro file.

At the root of the tree of profiles there is a group of profiles .
Each group of profiles can hold more than one evidence profile .
Each evidence profile contains in turn a set of relevant outcomes .

In a tree pane you may see, browse, rearrange, copy, move, and delete all profiles and outcomes stored in a given GRADEpro file.



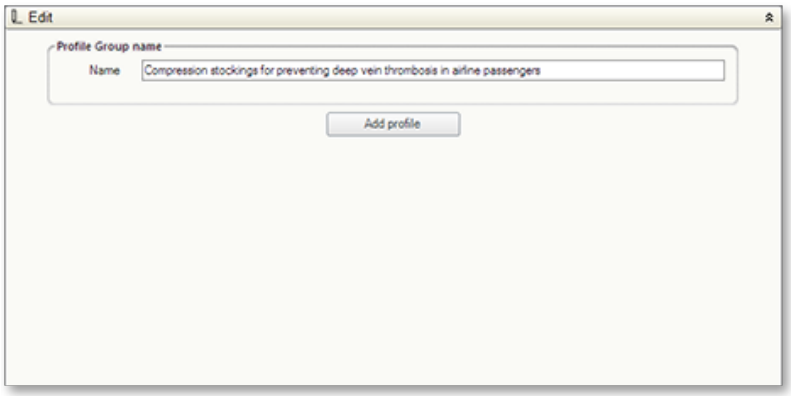
4.3.4 Edit pane

About Input and Edit Pane

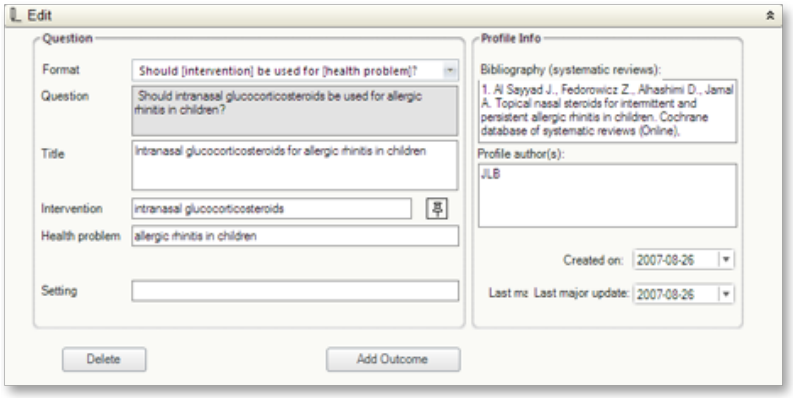
The input and edit pane provides a space in the GRADEpro interface where you **enter and edit** most of the data. Currently there are five layouts of editing panes in which you may:

- enter the name of profile group
- provide information about the profile: health care question and bibliography
- specify an outcome and rate its importance and quality of available evidence
- enter summary of findings for dichotomous outcomes
- enter summary of findings for continuous outcomes

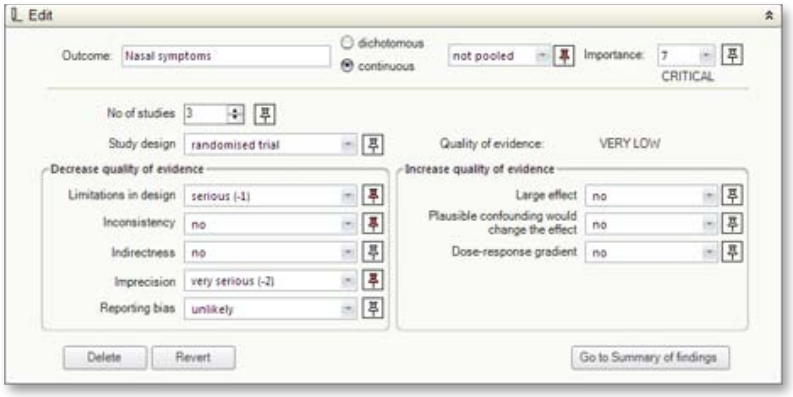
Name of profile group pane



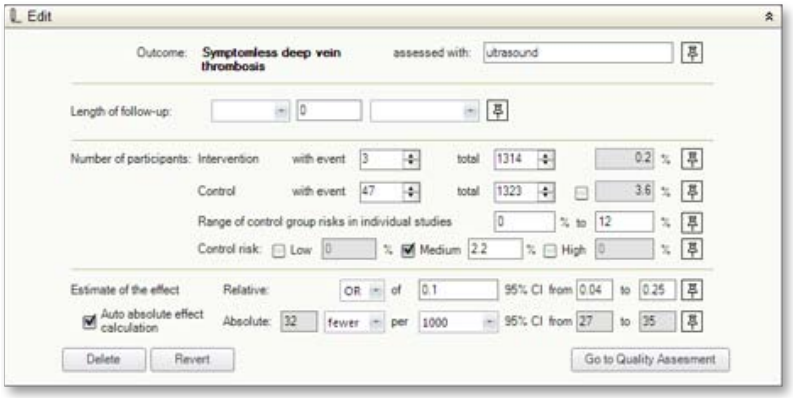
Information about the profile pane



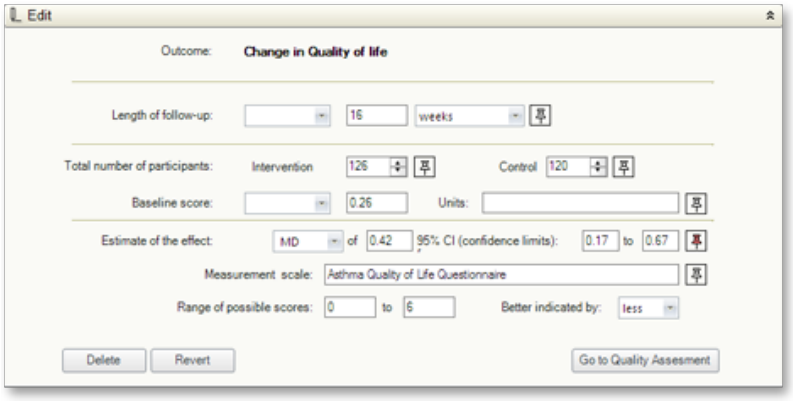
Outcome name, importance, and quality of evidence pane



Summary of findings for dichotomous outcome pane



Summary of findings for continuous outcome pane



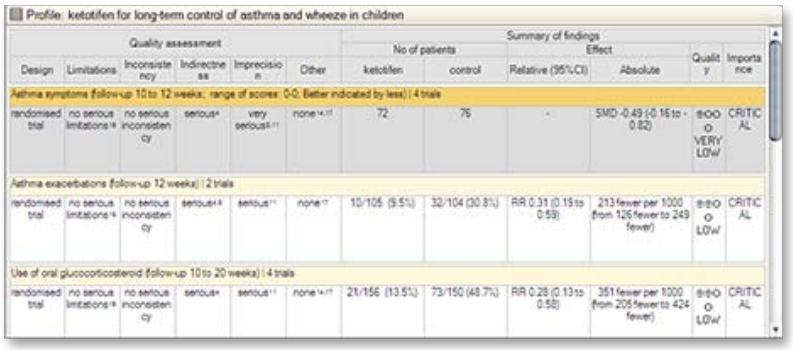
4.3.5 Preview of evidence table pane

About the Preview of Evidence Table Pane

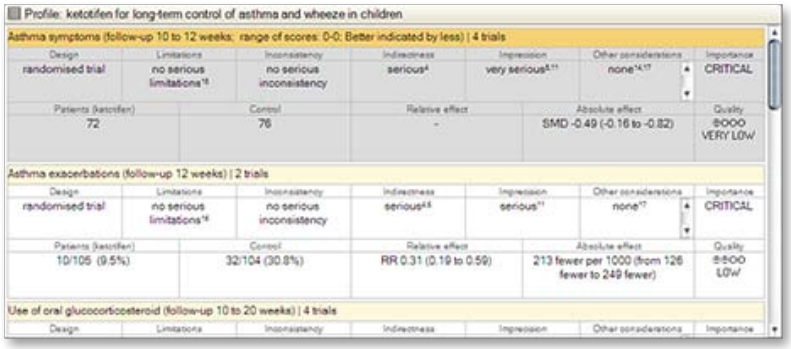
This section of the GRADEpro interface allows previewing how the information will be displayed in the evidence table you create. This preview will change depending on the type of evidence table you choose.

GRADE evidence profile

Single row presentation (a condensed view with each outcome in one row)

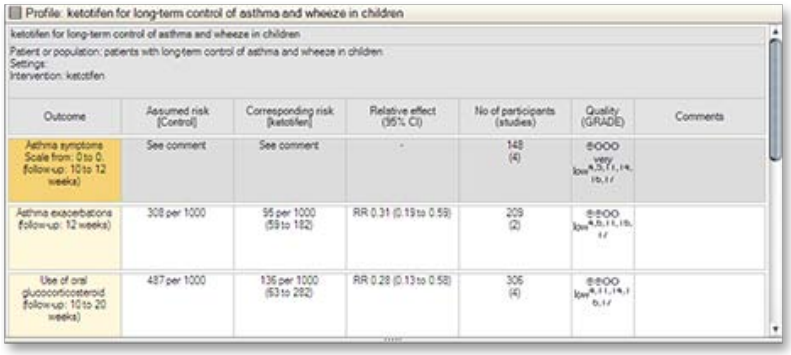


Two rows presentation (a broader view with the upper row showing the quality of evidence and lower one showing the summary of findings)



Summary of Findings (SoF) table

An on-screen depiction of a SoF table (note that the printed layout is graphically different)



4.3.6 Footnotes

About Footnotes

Footnotes allow **adding optional text** to any field in the evidence table. Typically you will use footnotes to **clarify information** you entered or **explain choices** you made when, for instance, grading the quality of evidence.

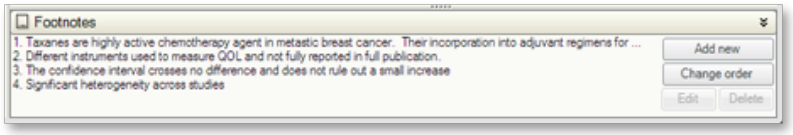
You may add a footnote to most fields in the evidence table at any time by clicking on a thumbtack. You may choose if to add a footnote or not, however, in certain circumstances you will be prompted to add it. A field to which a footnoted is assigned will be marked with salmon-colored thumbtack.

A footnote must be connected to at least one specific text or number entry box to appear in the evidence table. **One footnote can be connected to more than one entry**, such that the same footnote can be referred to repeatedly within a table. You may also **assign more than one footnote to any cell** in the evidence table. To learn more read about managing footnotes.

In the evidence (SoF) table, the footnote number will appear as a superscript above the text you entered in the entry box, and the footnote number and text of the footnote will appear at the end of the table in the footnotes section. The footnote number and text will be displayed in the order in which footnotes appear in the table. To learn more read about ordering and reordering

footnotes.

Footnotes are shown in **Footnotes pane**



4.3.6.1 Footnote order

About Footnote Order

Each footnote is numbered in Footnote manager. All footnotes that are checked for a given text box entry will show in the corresponding cell of the evidence table.

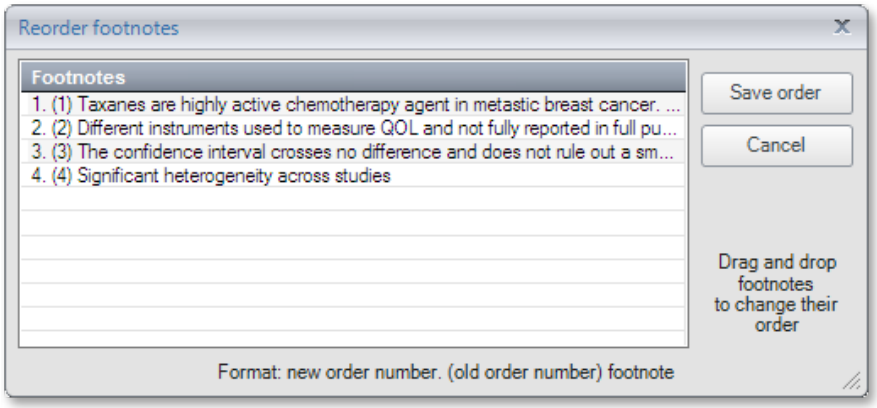
GRADEpro handles the order of footnotes differently depending on where you are viewing the footnotes:

- Footnote pane and Preview Evidence Table pane (**manual reordering**)
- Summary of Findings Preview and export (**automatic reordering**)

Footnote pane and Preview Evidence Table pane (manual reordering)

The footnotes shown in the Footnote pane and the Preview Evidence Table pane are ordered in the **sequence in which they were entered**. This order **does NOT correspond to the numbering and footnote order that will be shown in the exported evidence table** (Summary of Findings table and GRADE evidence profile).
You do not have to manually reorder footnotes since **all footnotes will be automatically reordered during export** to any format.

If you want to manually reorder footnotes in the Preview Evidence Table pane and in the Footnotes pane, so you see the footnotes in order of their appearance in the table (not the sequence of inserting), you can access **Change order** function in the Footnote pane or in the Footnote manager.



Drag and drop the footnotes in the preferred order. Then **save order**.

Summary of Findings Preview and export (automatic reordering)

During export or previewing the Summary of Findings table or GRADE evidence profile (click **Preview SoF** button on toolbar) **footnotes are automatically rearranged to reflect the order of their appearance** in the exported table: starting from top to bottom and from left to right.

Note

This feature will likely be enhanced in following versions of GRADEpro.

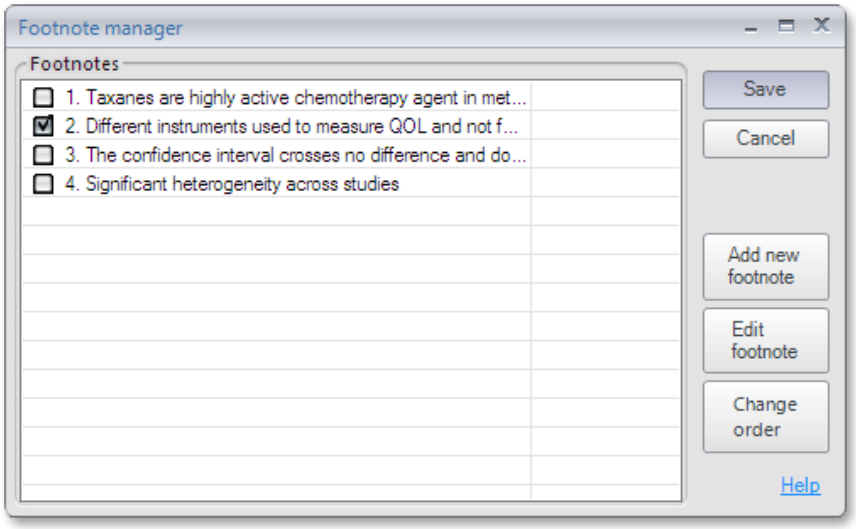
4.3.6.2 Footnote management

About Footnote Management

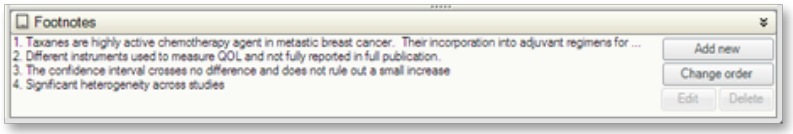
After footnotes have been entered, they can be managed using

- Footnote manager
- Footnote pane

Footnote manager



Footnote pane



Entering and Assigning footnotes

Once you have filled the information in the entry boxes in the Input and Editing pane, you can click on thumbtack located beside the text box and enter a footnote. A pop-up panel — **Footnote Manager** — will appear to help you manage all of the footnotes included in a profile. This will allow you to add a new footnote or select other previously entered footnotes when the footnote text is the same. Footnote manager includes **all footnotes entered for a single profile**.

To add a footnote in the Footnote Manager, click **Add footnote**, type in the text and then **save**. This footnote will automatically be checked to be included beside that text in the evidence table. Click **Save** to save changes to Footnote manager.

If you do not need to add a new footnote but you would like to **refer to a footnote that has been previously entered** and used, **check the box next to that footnote** in Footnote manager and save.

Note
Footnotes can also be added in the Footnote pane by clicking on **Add new**. However, any footnote that is added in this way will not be linked to any field in the table. To link that footnote to the text in the table, you must click the thumbtack in the Edit pane next to the field you want to assign a footnote to.

Editing footnotes

Footnotes can also be edited by clicking **Edit footnote** in the Footnote manager or the Footnote pane. Revise the text and then save.

Deleting footnotes

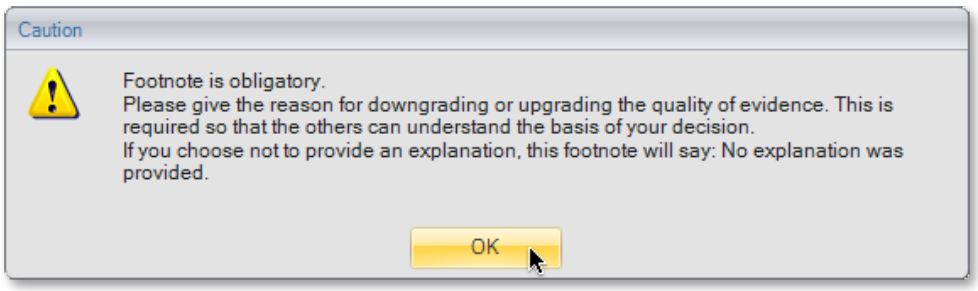
Footnotes can be deleted **only in the Footnote pane**. GRADEpro will ask you to confirm deletion. Once a footnote is deleted, GRADEpro will automatically update the order of footnotes.

4.3.6.3 Footnote prompt

About Footnote Prompt

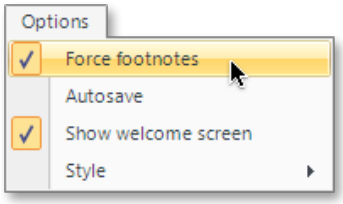
In certain circumstances you will be **prompted to add a footnote**.

The great merit of GRADE approach is **transparency** of decisions made when selecting the evidence to present in evidence tables and when grading the quality of the evidence. For this reason, whenever you downgrade or upgrade the quality evidence, or you do not downgrade/upgrade but want to make sure users know why, you are encouraged and will be prompted to add a footnote to explain your decision.



Once you click **OK** in the pop up box you can add or assign an existing footnote to this field.

You may disable or enable this function by deselecting or selecting **Force footnotes** in the **Options** menu. However we discourage switching it off to maintain transparency of your decisions.

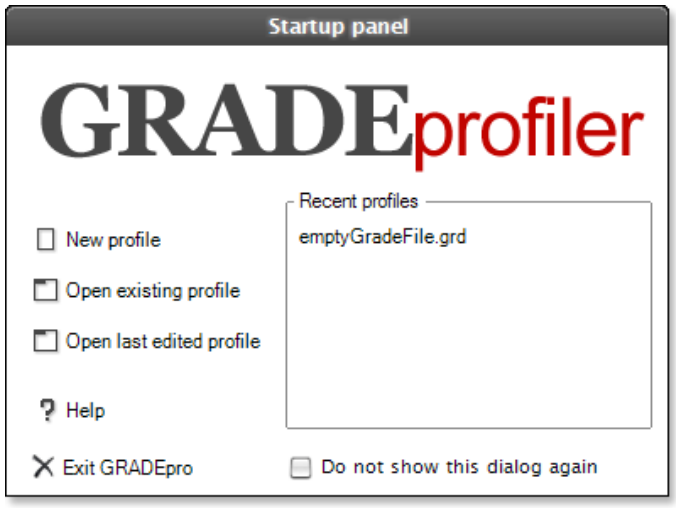


4.4 Welcome screen

About Welcome Screen

The welcome screen facilitates beginning work. It lets you choose what you do first:

- get help and read tutorial on how to use GRADEpro and the GRADE system — ?
- create a new evidence profile — **New profile**
- open an already saved evidence profile — **Open existing profile**
- open the **last edited or viewed profile**
- open one of **recently edited or viewed profiles**

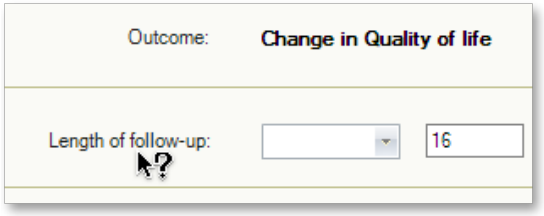


You can choose not to start GRADEpro with the Welcome screen in the future by checking **Don't show again**.
You can toggle the Welcome screen on/off choosing **Options » Show welcome screen**.

4.5 Context-specific help

About Context-specific Help

For many topics GRADEprofiler provides a context-specific help. When you **hover** with your mouse pointer over the "title" of each field, a cursor changes into one with question mark . If you click it takes you to the help section on the particular topic.



4.6 Keyboard shortcuts

Keyboard shortcuts

You can use the following keyboard shortcuts while working with GRADEprofiler

SHORTCUT	ACTION
Ctrl+O	Open
Ctrl+P	Print
Ctrl+Q	Quit
Ctrl+S	Save
Ctrl+Shift+G	Add new Group of Profiles
Ctrl+Shift+O	Add new Outcome
Ctrl+Shift+P	Add new Profile
F1	Help
Tab	moves cursor to next field
Shift+Tab	moves cursor to previous field

5. Creating GRADE evidence profiles

Creating GRADE Evidence Profiles

You can use GRADEprofiler to create a GRADE evidence profile. A GRADE evidence profile allows presentation of key information about all relevant outcomes for a given health care question focusing on three key parts:

- information about the health care question defining a profile
- grade of the quality of the evidence
- summary of the statistical results

You will need to enter all data manually for each outcome or import and then edit data from a Review Manager 5.0 file. When GRADE evidence profile is finished, you can save the GRADEpro file for later use, export it or print.

Overall tasks to create a GRADE evidence profile

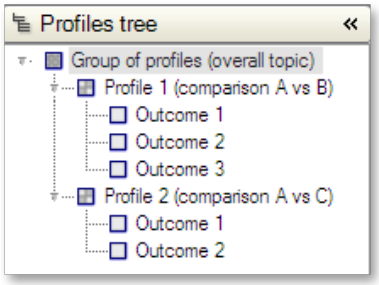
1. Open GRADEpro.
2. Choose **New Profile** in the welcome screen and name the file.
3. Create profile group
4. Create a profile (specific health care question)
5. Add or edit the information about the comparison (health care question and information about bibliography and authors).
6. Create an outcome
7. Add, move or delete outcomes in each profile.
8. Select an outcome. For each outcome there are 2 sections: Summary of Findings screen and the Quality Assessment screen.
9. Select the Quality Assessment screen. Complete it by first confirming number and type of studies, then assess the quality of evidence for the outcome. Downgrade or upgrade evidence according to GRADE criteria and enter footnotes when necessary.
10. Select Summary of Findings screen. Add data.
11. Repeat (#9 and #10) for all outcomes in the profile.
12. If, there is more than 1 comparison of importance repeat (#4 to #11) to create another profiles.
13. Preview GRADE evidence profile, double check presentation and edit if necessary. **Do not skip this step.**

5.1 Creating a group of profiles

Creating a Group of Profiles

Groups of profiles represent an **overall topic** of a systematic review or overall topic of a health care recommendation, that may involve consideration of several health care questions (comparisons presented in evidence profiles) or an overview of reviews.

Any GRADEprofiler file has to contain at least one group of profiles at the root of the tree of profiles.



The **group of profiles** allows organizing several evidence profiles regarding a similar topic in one batch. It facilitates viewing, organizing, and saving evidence profiles for similar health care questions. It also allows exporting Overview of Reviews table.

You may add an additional group of profiles within the same GRADEprofiler file whenever you want to put together evidence tables containing information on another health care problem.

To create a Group of Profiles:

- Choose **Profile Group** from the **Add** menu or click on a toolbar.
- Type the name of a Profile Group.
- Click **Add profile**.

A new **Add Profile** dialog box will automatically appear.

Example

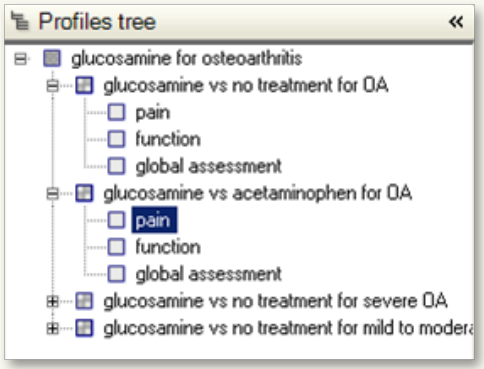
You are interested in creating several evidence profiles concerning different management options for lung cancer. One Group of Profiles could contain health care questions about surgical methods and the other about medical management of lung cancer.

For authors of systematic reviews

In GRADEpro, a systematic review will usually be organised into three levels:

- **Profile Group**: the main topic of the review. For example, glucosamine and other interventions for osteoarthritis.
- **Profiles**: the (multiple) comparisons in the review. For example, glucosamine versus no treatment, and glucosamine versus acetaminophen OR glucosamine versus no treatment in severe osteoarthritis and glucosamine versus no treatment in mild to moderate osteoarthritis. Since a review may have more than one comparison and also comparisons with subgroup analyses, there will often be more than one profile in one review. Each profile is set up as a more specific or focused health care question than the Profile Group.
- **Outcomes**: the outcomes measured in each comparison. For example, pain, function, global assessment, and adverse effects.

The tree for the example above would appear in GRADEpro as



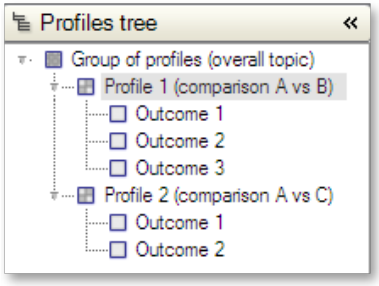
For authors that import data from RevMan the profile group, profiles, and outcomes will be imported. It may be necessary to move, delete or add other profile groups, profiles, and outcomes depending on how the comparisons and outcomes were created and organised originally in RevMan.

5.2 Creating a profile

Creating a Profile

Evidence profile (GRADE evidence profile or Cochrane Summary of Findings table) is a table summarizing the information about the outcomes for a focused health care question. It provides simple and transparent summary of available evidence about a particular health care management option.

A profile is a subdivision of a profile group . An evidence profile is defined by a specific health care question following the PICO format. It **represents a single comparison** (*e.g.* intervention versus placebo) for which multiple outcomes may have been measured.



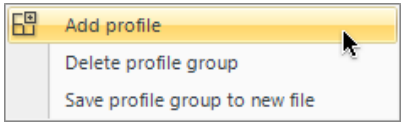
To create a Profile:

- Choose **Profile** from **Add** menu.
- Specify a health care question.
- Type or paste bibliographic information.
- Type the name of the author of the profile.

Note

There are four possibilities to create a new Profile:

- Choose **Profile** from the **Add** menu
- Click on a toolbar
- Press **Ctrl + Shift + P**
- Right-click on a Group of Profiles or on a Profile in the tree on the left-hand side and choose **Add Profile**



5.2.1 Defining a health care question

Defining a Health Care Question

You should define a health care question whenever you create a new evidence profile.

We recommend that every comparison in a systematic review or a recommendation in guidelines is defined by a clear, explicit, and focused health care question. The clearer the question the easier it is to formulate an unambiguous conclusion or recommendation.

To formulate a question:

- choose question format from a drop-down menu
- specify the **intervention** being considered
- specify the alternative intervention (**comparison**)
- depending on the question format you have chosen, specify **patients or population** that the recommendation is intended for, or a **health problem** that the recommendation will address

You may also specify the **setting** from which the evidence is obtained (*e.g.* outpatient vs. inpatient, the countries in which the trials were performed etc.).

> *For authors of systematic reviews*
>
> Health care question and other information in this editing pane is the **content for the top descriptive section of the Summary of Finding table**. Authors who have imported data from a Review Manager File will note that much of the information has been imported, but will likely need to **edit the information**.
>
> Authors of systematic reviews may specify **setting** in which reviewed studies were done (*e.g.* developing vs developed countries, inpatient vs outpatient, etc.).

» more about health care question

5.2.2 Providing profile information

Providing Profile Information

The profile information pane allows you to:

- specify **references** to any systematic reviews or other sources of evidence you used to create a profile
- provide the name of the **person who created the profile** (most often it is you)
- specify the **date when the profile was updated**.

This information you usually enter once when you create a new profile, although you may wish to update bibliography or information on the date you last changed the profile.

When entering the information on references you may either **type** them or **copy and paste** them from another document.

Tip for authors of systematic reviews

This part of evidence profile will **not be displayed in the Summary of Findings table**, so it is not necessary to provide this information when your only goal is to create SoF table to export to RevMan.
Part of the bibliographic information is inserted automatically when importing data from Review Manager 5.0 file.

5.3 Creating an outcome

Creating and Managing Outcomes

You can specify one or more outcomes for any health care question indicated in the profile.

In most clinical situations patients are interested in many outcomes (*e.g.* patients with asthma may be interested in preventing an exacerbation requiring hospitalization, need for administration of oral glucocorticosteroid, aggravation of symptoms disturbing their sleep, etc.).

It is important to include in the evidence table (Summary of Findings table or GRADE evidence profile) **all outcomes that are ultimately important to patients** regardless if they were measured in the studies or not. Ideally authors of a systematic review would **specify all patient-important outcomes before conducting a review** and those making recommendations should specify these outcomes at the beginning of guideline development process. Outcomes that were either not reported, not measured, or not pooled, but considered important to patients and would affect decision making, should still be included in the evidence table (SoF table), although one can neither grade quality of supporting evidence (because it is not available) nor provide the estimated magnitude of effect. For these outcomes the ratings are left blank (see examples below).

SoF table

A compared to B for some health problem

Patient or population: patients with some health problem
Settings:
Intervention: A
Comparison: B

Outcomes	Illustrative comparative risks* (95% CI) Assumed risk B	Corresponding risk A	Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
Quality of life - not measured	See comment	See comment	Not estimable	-	See comment	Quality of life was not measured it any of the studies

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval;

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

GRADE evidence profile

Question: Should A vs B be used for some health problem?
Settings:
Bibliography:

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	A	B	Relative (95% CI)	Absolute		
Quality of life - not measured												
0	-	-	-	-	-	none	0	0	-	-		CRITICAL

Information about the outcome

Each outcome in the evidence table (SoF table) consists of two sets of data: quality of the evidence and the results of the studies. There are two panes in GRADEpro to enter and edit that data:

1. quality assessment of the evidence
2. summary of findings or the results

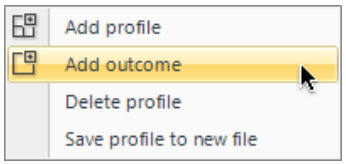
You will need to add, edit, or delete information in each edit pane to fill in the evidence (SoF) table.

To create an Outcome:

- Choose **Outcome** from **Add** menu.
- Name the outcome.
- Choose the type of the outcome (dichotomous or continuous).
- Rate importance of the outcome.

There are four possibilities to create a new Outcome:

- Choose **Outcome** from **Add** menu
- Click on a toolbar
- Press **Ctrl + Shift + O**
- Right-click on a Profile in the tree on the left-hand side and choose **Add Outcome**



For authors of systematic reviews

You will likely not include all of the outcomes from the review in the Summary of Findings table (a **maximum of 7 outcomes are recommended**).

When importing data from a Review Manager 5.0 file

1. All outcomes that are summarized in with a **meta-analysis** (including subgroups) will be imported into GRADEpro (see below for exceptions).
2. Outcomes with **sub-totals only** will not be imported. However, the subgroup outcomes if totalled will be imported.
3. Outcomes with no totals are imported. Authors may want to choose 'single study' when defining the status of the effect measure of an outcome.
4. Authors will need to move, add or delete outcomes depending on which outcomes were imported and which outcomes are important to patients. You can order the outcomes by their importance by draging them in the tree pane.

5. Imported from the Review Manager 5.0 file are: information about the type of outcome, number of studies and some of the data from the meta-analysis such as control and intervention events, point estimates. Other data will need to be entered (see subsections under Completing Quality Assessment for an outcome and Completing the Summary of Findings for an outcome).

5.3.1 Entering outcome information

Entering Basic Information About an Outcome

In this row you can enter the name of an outcome, specify its type and status, and rate its relative importance. This information will help to determine if the outcome should be included in the evidence table and the way it is presented.

You will enter this information for each outcome.

- Name an outcome
- Choose the type of an outcome (dichotomous or continuous)
- Choose the status of the analysis for this outcome
- Rate relative importance of an outcome

For authors of systematic reviews

Name of the outcome and its type are inserted automatically when importing data for Review Manager 5 file. If the outcome is imported its status is set to **pooled**. However, you always have to rate the relative importance of the outcome in GRADEpro — it is not rated automatically.

5.3.2 Types of outcomes

About Types of Outcomes

There are two types of outcomes: **dichotomous** and **continuous**, that require different presentation of findings, because their summary statistics differ.

Dichotomous (binary) outcomes

Each individual's outcome may have one of only two possible categorical responses (e.g. dead or alive, myocardial infarction or no myocardial infarction, etc.).

Continuous outcomes

The term 'continuous' in statistics conventionally refers to data that can take any value in a specified range. A continuous outcome may be either:

- "truly" **continuous** (e.g. blood glucose concentration or weight)
- **ordinal** (e.g. score on a Hamilton depression scale or Chronic Respiratory Questionnaire)

Summary statistic and the presentation of results for both types of "continuous" outcomes are usually the same.

> *For authors of systematic reviews*
>
> The type of an outcome will determine the data that you will need to enter in GRADEpro. For authors who import data from a Review Manager 5.0 file, the type of outcome will be determined automatically.

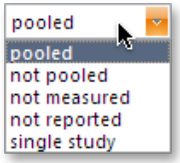
5.3.3 Specifying Status of the Analysis for an Outcome

Specifying the Status of the Analysis for an Outcome

Status of the analysis for an outcome refers to one of a number of possibilities **how an effect measure is be presented**.

You can specify the status of a particular analysis for an outcome as:

- **pooled** — the outcome was reported in at least 2 studies and a **meta-analysis was conducted** and there is a combined estimate of effect
- **not pooled** — the outcome was reported in at least 2 studies but a **meta-analysis was not conducted** and there is no combined estimate of effect (*e.g.* due to heterogeneity); in this case you may want to give a range of the results of studies
- **not measured** — the outcome was not measured in individual studies
- **not reported** — individual studies may indicate, that the outcome was measured but they did not report the results
- **single study** — there is only one study available (GRADEpro handles this status similar to "pooled", however this category was separated for logical resons (one cannot pool the results of a single study)



Note
It is also important to **provide an explanation** in a footnote or comment describing why studies were not pooled.

> *For authors of systematic reviews*
>
> When the data are imported from a Review Manager 5 file all imported outcomes have a status of being **pooled**.

5.3.4 Rating importance

Rating Importance of the Outcome

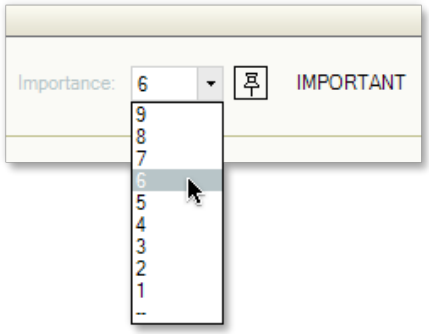
The GRADE approach suggests rating the importance of each outcome on a **9-point scale** from 1 to 9.

A rating or assessment of the importance of outcomes is necessary to choose which outcomes should be considered in deciding about the benefits and downsides of an intervention or about which outcomes should be included in a GRADE evidence profile or Summary of Findings (SoF) table.

Guideline panels and other users making decisions about the importance of outcomes are encouraged to rate the importance of each (newly **added**) outcome and **reconsider** it after completing the summary of findings for this outcome. Usually the importance of an outcome should not be influenced by the statistical results. However, there are exceptions when, for instance, after compiling the evidence an outcome is found to be extremely rare. In such situations reconsideration of the importance may be necessary.

To rate the importance of an outcome:

- Choose a number from 1 to 9 from a drop-down list
- Explain your choice in a footnote if you think it may be imprtant for the others to understand your choice.



5.4 Grading quality of evidence

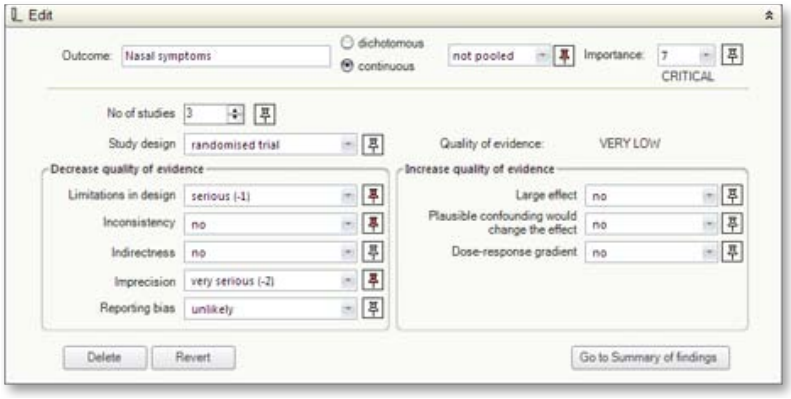
Grading Quality of Evidence

You should **assess the quality of the evidence for each outcome**, with the **exception** of outcomes for which there are no studies that evaluated it (**not measured**) or no studies reported this outcome (**not reported**).

The editing pane is divided into:

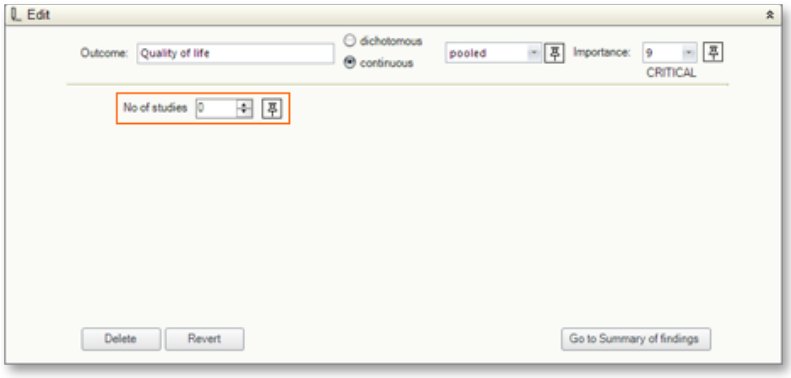
- information about the outcome
- quality assessment
  - number of studies
  - type of studies

- factors that affect the quality of the evidence

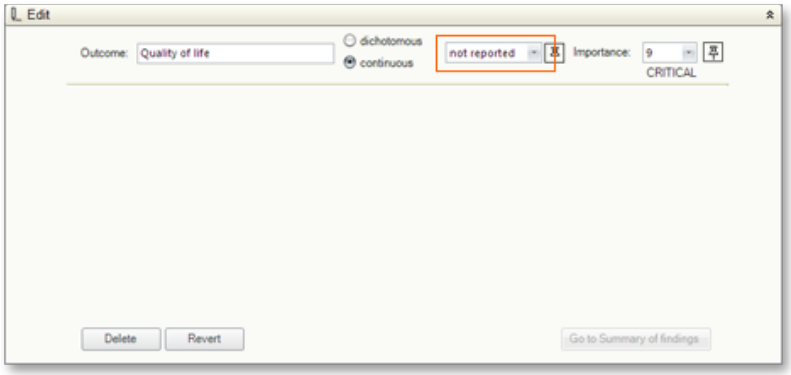


Note

The editing pane for Quality Assessment will appear **empty until you choose at least one study and type of studies** contributing to the results.



It will also be empty when the outcome was either not measured or not reported. In this case you will not be able to proceed to summary of findings pane since there are no findings to report (however, a **Summary of Findings (SoF) table can be generated**).



5.4.1 Specifying the number of studies

Reporting the Number of Studies

You should specify the total number of studies included in the systematic review that examined this outcome.

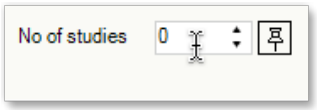
Specifying the number of studies informs users whether the outcome was addressed in:

- no study — there are no published data that fulfilled the inclusion criteria for the question asked
- only in one study
- many studies.

You should specify the number of studies everytime you add a new outcome.

To specify the number of studies:

- Type a natural number or use the up and down arrows in a text box
- Explain your choice in a footnote if you think it may not be clear why you have chosen that number.



For authors of systematic reviews

When importing data from a Review Manager 5.0 file, the number of studies that were included in the meta-analysis for the given comparison for a specific outcome will be automatically entered here.

5.4.2 Choosing study design

Choosing study design

You can specify the basic design of studies included in the systematic review that examined this outcome. Choose a basic study design that is an essential determinant of the quality of evidence:

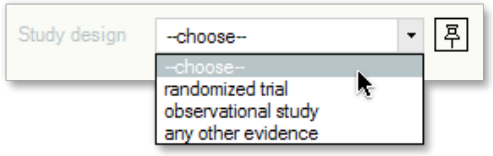
- randomised trial
- observational study
- any other evidence (case series or case reports)

Note
The category referring to **any other evidence** such as case series or case reports is currently included, but users may choose to include this category in observational studies and downgrade in the detailed assessment of study design and limitations for the absence of an independent control group.

Choosing the basic study design is a prerequisite to grading the quality of evidence for any outcome.

To choose the basic study design:

- Choose a study design from a drop-down list
- Add any other information in a footnote if you think it may be important for others to know this.



5.4.3 Quality of evidence display

About the Quality of Evidence Preview

The preview of the quality of evidence grade will be displayed only **after you finish the grading**, *i.e.* after you rate all factors that influence it.

5.4.4 Rating factors that influence quality

Rating Factors that Influence Quality of Evidence

You should carefully consider and rate all the factors that may influence the quality of evidence.

You should carefully consider and rate all the factors that may influence the quality of evidence. While doing so, you should bear in mind that down- and upgrading for specific quality factors should be done in the context of all of the factors that influence the quality of evidence. Having downgraded for one quality criterion may influence how the next quality criterion is dealt with. Thus, it could influence the threshold for downgrading of another criterion. While it is acceptable to avoid judgments that might be overly severe, the decisions should be made transparent. For example, borderline situation in which it would be reasonable either to downgrade for study limitations, or not to do so may exist. This illustrates that the great merit of GRADE is not that it ensures consistency of conclusions, but that it requires explicit and transparent judgments. In any case you should explain your decision in a footnote.

Factors that influence Quality of Evidence

Decrease	Increase
Study limitations	Large magnitude of effect
Inconsistency of results	All plausible confounding would reduce the demonstrated effect
Indirectness of evidence	Dose-response gradient
Imprecision	

Publication bias

Related topics

- Detailed description of factors influencing quality of evidence
- Diagram for rating the quality of evidence in GRADE system

5.4.4.1 Rating study limitations

Rating Limitations in Study Design or Execution (Risk of Bias)

You should assess if the studies had limitations in design or execution that were serious enough to downgrade the quality of evidence for this outcome.

To rate study limitations:

- If you think any limitations were **negligible** choose **no**
- If you think there were **serious** limitations choose **serious**
  «this will downgrade the quality of evidence for this outcome by **1 level**»
- If you think there were **very serious** limitations choose **very serious**
  «this will downgrade the quality of evidence for this outcome by **2 levels**»
- Explain your choice in a footnote whenever you downgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Authors of Cochrane Collaboration systematic reviews, can use tools provided in chapter 8 of the Cochrane Handbook to assess risk of bias. It provides a detailed discussion of study-level assessments of risk of bias in the context of a Cochrane review and proposes an approach to assessing the risk of bias for an outcome across studies as »low risk of bias«, »unclear risk of bias« and »high risk of bias«. These assessments may be used directly to inform the assessment of study limitations in the GRADE approach. In particular:

- **low risk of bias** would indicate **"no serious limitations"**
- **unclear risk of bias** would indicate either **"no serious limitations"** or **"serious limitations"**
- **high risk of bias** would indicate either **"serious limitations"** or **"very serious limitations"**.

Authors of Cochrane systematic review must use their judgment to decide between alternative categories, depending on the likely magnitude of the potential biases.

Related topics

- Detailed description of study limitations in the GRADE system
- Chapter 8 of the Cochrane Handbook "Assessing risk of bias in included studies"

5.4.4.2 Rating inconsistency

Rating Inconsistency in Results

You should assess if the results were consistent across studies and if any inconsistency may have been serious enough to downgrade the quality of evidence for this outcome.

To rate inconsistency:

- If you think any inconsistency was **negligible** choose **no**
- If you think there was **serious** inconsistency choose **serious**
  «this will downgrade the quality of evidence for this outcome by **1 level**»
- If you think there was **very serious** inconsistency choose **very serious**
  «this will downgrade the quality of evidence for this outcome by **2 levels**»
- Explain your choice in a footnote whenever you downgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Related topics

» Detailed description of inconsistency in the GRADE system

5.4.4.3 Rating indirectness

Rating Indirectness of Evidence

You should assess if the evidence answers directly the health care question you have asked and if any indirectness of available evidence may have been serious enough to downgrade the quality of evidence for this outcome.

To rate indirectness:

- If you think the evidence is direct choose **no**
- If you have **serious** doubts about directness choose **serious**
  «this will downgrade the evidence for this outcome by **1 level**»
- If you have **very serious** doubts about directness choose **very serious**
  «this will downgrade the evidence for this outcome by **2 levels**»
- Explain your choice in a footnote whenever you downgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Related topics

» Detailed description of indirectness in the GRADE system

5.4.4.4 Rating imprecision

Rating Imprecision of Results

You should assess if the results are precise enough and if any imprecision of the results may have been serious enough to downgrade the quality of evidence for this outcome. Imprecision is

defined differently for authors of systematic reviews and for guideline panels.

To rate imprecision:

- If you think the results were precise choose **no**
- If there was **serious** imprecision choose **serious**
  «this will downgrade the quality of evidence for this outcome by **1 level**»
- If there was **very serious** imprecision choose **very serious**
  «this will downgrade the quality of evidence for this outcome by **2 levels**»
- Explain your choice in a footnote whenever you downgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Related topics

» Detailed description of imprecision in the GRADE system

5.4.4.5 Rating publication bias

Rating Publication Bias

You should assess if there is a probability of publication bias and if reporting bias may have been serious enough to downgrade the quality of evidence for this outcome.

To rate probability of the publication bias:

- If you think there is no evidence of publication bias choose **unlikely**
- If there is **high** probability of publication bias choose **likely**
  «this will downgrade the quality of evidence for this outcome by **1 level**»
- If there is **very high** probability of publication bias choose **very likely**
  «this will downgrade the quality of evidence for this outcome by **2 levels**»
- Explain your choice in a footnote whenever you downgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Related topics

» Detailed description of publication bias in the GRADE system

5.4.4.6 Rating magnitude of effect

Rating Magnitude of the Effect

You should assess if the effect was large or very large and, if so, upgrade the quality of evidence accordingly for this outcome.

To rate magnitude of the effect:

- If the effect was not large (RR between 0.5 and 2.0) choose **no**
- If the effect was **large** (RR either >2.0 or <0.5 based on consistent evidence from at least 2 studies, with no plausible confounders) choose **RR >2 or <0.5**
  «this will upgrade the quality of evidence for this outcome by **1 level**»
- If the effect was **very large** (RR either >5.0 or <0.2 based on direct evidence with no major threats to validity) choose **RR >5 or <0.2**
  «this will upgrade the quality of evidence for this outcome by **2 levels**»
- Explain your choice in a footnote whenever you upgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Note
A relative risk (RR) of 0.5 corresponds to a relative risk reduction (RRR) of 50% and a RR of 0.2 corresponds to a RRR of 80%.
When the events in the control group are not frequent, OR and HR can be assumed to be equal to the RR for the application of this criterion.

Related topics

» Detailed description of the magnitude of the effect in the GRADE system

5.4.4.7 Rating dose-response gradient

Rating the Dose-Response Gradient

In **randomised trials** and in **observational studies downgraded for any reason** do not rate the presence of dose-response gradient and choose **no**

You should assess if there was a dose-response gradient only in **observational studies not downgraded for any reason**. If a dose-response gradient was present, upgrade the quality of evidence for this outcome.

To rate the presence of dose-response gradient:

- If there is no evidence of dose-response gradient choose **no**
- If there is evidence of dose-response gradient choose **yes**
  «this will upgrade the quality of evidence for this outcome by **1 level**»
- Explain your choice in a footnote whenever you upgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Related topics

» Detailed description of the dose-response gradient in the GRADE system

5.4.4.8 Rating influence of all residual confounding

Rating the Influence of All Plausible Residual Confounding

In **randomised trials** and in **observational studies downgraded for any reason** do not rate the influence of all plausible residual confounding and choose **no**

Only in **observational studies not downgraded for any reason** you should assess if the influence of all plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results show no effect. In either of these two cases, upgrade the quality of evidence for this outcome.

To rate the effect of all plausible residual confounding:

- If there is no evidence that the influence of all plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results show no effect choose **no**
- If there is evidence that the influence of all plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results show no effect choose **yes** «this will upgrade the quality of evidence for this outcome by **1 level**»
- Explain your choice in a footnote whenever you upgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Related topics

» Detailed description of the influence of all plausible residual confounding in the GRADE system

5.5 Summarizing the evidence

Completing Summary of Findings for an Outcome

Summary of findings contains the information about the **usual burden of the outcome** (control risk or control score) and the **magnitude and precision of the estimated effect**.

For each outcome, it is necessary to provide the results from the meta-analyses or single studies. These results are entered and presented differently depending on whether the outcome is dichotomous or continuous. The editing pane for the GRADEevidence profile, therefore, appears differently for each type of outcome.

Depending on whether the outcome is dichotomous or continuous, results are entered and presented differently and the editing pane appears differently for each type of these outcomes.

In general, the data necessary to complete the summary of findings include:

1. description of the outcome, how it was assessed and in what time period
2. number of participants in the studies
3. baseline risk either for dichotomous or continuous outcome
4. estimate of the relative effect for a dichotomous outcome or of the effect for continuous outcome
5. absolute effect for a dichotomous outcome

Detailed description for data entry and editing for

» dichotomous outcomes

» continuous outcomes

5.5.1 Dichotomous outcomes

Completing Information about Dichotomous Outcomes

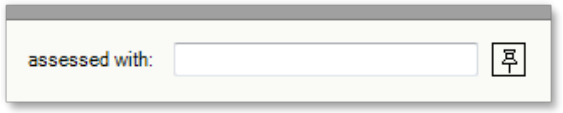
You will need to enter or edit (depending on whether data were imported from Reviw Manager 5.0 file) the following information for a dichotomous outcome:

- how the outcome was assessed
- length of follow-up
- number of participants
- illustrative risk on control intervention (baseline risk)
- relative effect
- absolute effect

5.5.1.1 Providing outcome measure

Providing Information on How the Outcome was Assessed

You may provide additional information on how the outcome was assessed by typing it in the box.



Certain outcomes may be assessed in different ways (e.g. one may use transesophageal echocardiography, computed tompgraphy, magnetic resonance imaging, or interventional angiography to assess the size of thoracic aotric aneurism). You may provide his information, if you belive it is important that guideline panels or anyone making decisions based on the information provided in the evidence profile knows the method in which the outcome was assessed.

5.5.1.2 Length of Follow-up

Specifying the Length of Follow-up

You can provide information on the length of the follow-up in the studies that reported each outcome. This is important, since the interpretation of the observed effect depends on the time in which the events occurred.

You will have to make **judgments across studies**, because it is unlikely that the duration of follow-up was the same in all studies.

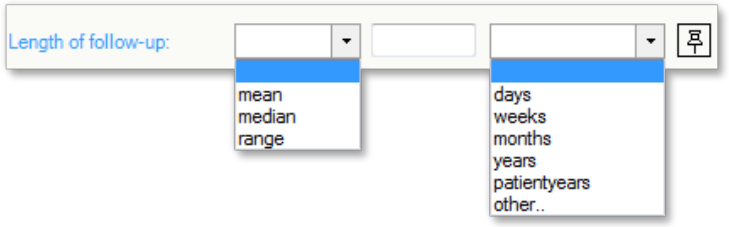
Note

The **duration of intervention** (which should be described in the intervention type, *e.g.* treatment with glucosamine for 8 weeks) may differ from the **duration of follow-up** (*e.g.* assessment of health related quality of life or function after 1 year).

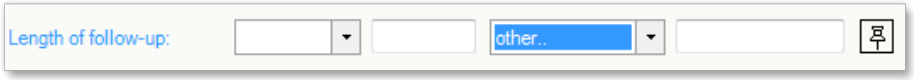
You should provide the information on the length of follow-up for each outcome separately since they may have been measured in different periods.

To provide the information on the time of follow-up:

- Choose **mean**, **median**, or **range** from a drop-down list
- Type a **number** (for mean or median) or a range (e.g. **4 to 7**) in the box
- Choose the unit of time from a drop-down list (**days**, **weeks**, **months**, **years**, **patient-years**, or **other...**)



If you choose other... measure of time an additional box will appear allowing you to specify the unit (*e.g.* hours, etc.)



For authors of systematic reviews

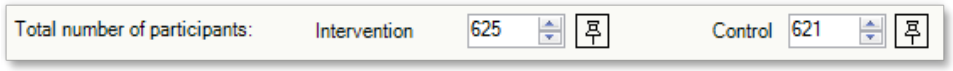
The length of follow-up at which time the outcome was measured will always need to be entered regardless of whether data were imported from a Review Manager 5.0 file or not.

5.5.1.3 Total number of participants

Providing Number of Participants

You will need to enter a total number of participants in control groups and intervention/treatment groups in the studies that contribute to the results. Usually this information can be obtained directly from a meta-analysis graph (forest plot).

For a **continuous outcome** you should specify **total number of participants** in the intervention/treatment and control groups.



For a **dichotomous outcome** you should also specify the **number of participants (patients) in whom the event occurred**.

In the **grey boxes to the right** you will see the **percentage** of total subjects in whom the outcome occurred. The check-box next to the control group event rate lets you choose this rate as a basis for calculation of the absolute effect observed in the studies.

> *For authors of systematic reviews*
>
> When **importing** from Review Manager 5.0 file this information will be **automatically entered** into GRADEpro.

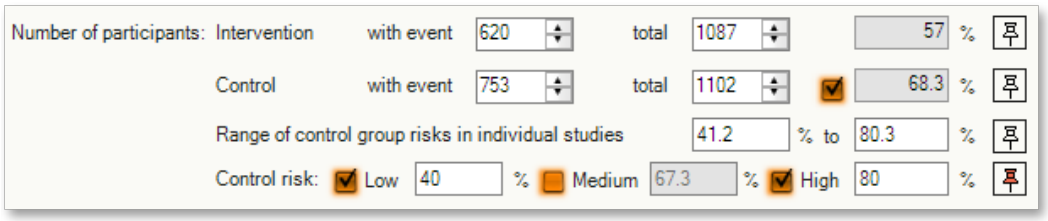
5.5.1.4 Control risk

About Control (Baseline) Risk

The results presented in the GRADE evidence profile are built around the assumption of a **consistent relative effect**. It is therefore important to consider the **implications of this effect for populations at different baseline risks**. An illustrative risk on control intervention (i.e. baseline risk or assumed risk) is a measure of the typical burden of the outcomes.

You can provide **up to three typical risk values (illustrative risk on control intervention or assumed baseline risk)** out of **four possibilities** provided in GRADEpro. There are four checkboxes (here colored orange for emphasis) next to fields where assumed risks may be entered.

Below you will find few suggestions how to choose illustrative control risk(s) and how to present it.



It is important to indicate with a footnote why a particular control risk was chosen.

In the above example three assumed control risks were provided: mean risk on control intervention in the studies included in the systematic review, low risk, and high risk. These will be shown in a GRADE evidence profile as three separate populations at different baseline risk and a corresponding absolute effect will be calculated for each of illustrative control risks based on the relative effect (presumed to be constant irrespective of baseline risk).

Question: Should bisphosphonates added to anti-cancer treatment be used in women with advanced breast cancer and bone metastases?
Settings: ???
Bibliography: Pavlakis N, Schmidt RL, Stockler MR. Bisphosphonates for breast cancer. Cochrane Database of Systematic Reviews [Year], Issue [Issue].

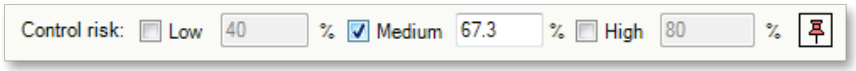
Quality assessment							Summary of findings					
							No of patients		Effect		Quality	Importance
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	bisphosphonates added to anti-cancer treatment[4]	control	Relative (95% CI)	Absolute		
Skeletal events (new bone metastases, fractures, spinal cord compression, irradiation or surgery, bone pain) (follow-up 1-3 years)												
8	randomised trial	no serious limitations	no serious inconsistency	serious[1]	no serious imprecision	none	620/1087	753/1102	RR 0.83 (0.78 to 0.89)	12 fewer per 100 (from 8 fewer to 15 fewer)	⊕⊕⊕O MODERATE	
								40%		6 fewer per 100		
								80%		13 fewer per 100		

Choosing the assumed control risk to present in the GRADE evidence profile

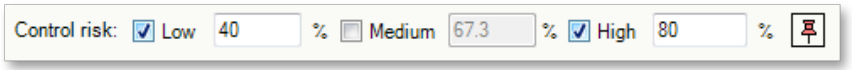
The assumed risk can be derived from a variety of populations at different baseline risk of an outcome and at different lengths of follow-up. Ideally, risks would reflect groups that clinicians can easily identify on the basis of their presenting features.

Suggestions:

- present the **median control group risk from the studies** included in a meta-analysis.
  If there is little variation in the baseline risks across the studies included in the meta-analysis, you may calculate the median control group risk across studies.



- present up to 3 risks based on the **control group risks in the studies** included in the meta-analysis. You can calculate a low, medium and high assumed risk from the studies. Alternatively, for a high and low risk population you can choose the **second highest** and **second lowest** control group risks in the included studies.

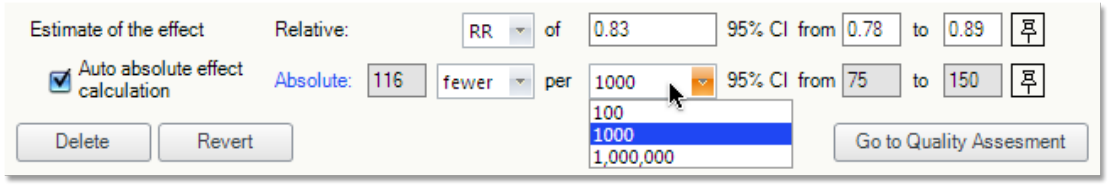


- present a baseline risk **from observational studies**. You may enter it as a low, medium, or high risk depending on your judgement, and explain where the assumed risk values come from.
- present a **mean baseline risk from the studies** included in meta-analysis calculated as number of patients in the control groups with event divided by a total number of patients in control groups. This is automatically calculated by GRADEpro and displayed after the total number of subjects in control groups. This is a natural choice when there is only one study available.

Whichever assumed risk you choose to present you should explain your choice in a footnote.

Choosing the denominator to present risks

You can present assumed risk and absolute effect of an intervention as number of subjects per 100, per 1000, or per 1,000,000. We suggest that, **by default, risk is presented per 1000 people**. You can choose to present risk per 1,000,000 if the events are rare or per 100 if the events are more frequent.



Providing range of control-group risks in individual studies

Information about the range of control-group risks in individual studies is provided because it allows:

- choosing the control group risk(s) that will appear in the evidence profile (it serves as a step in completion of an evidence profile, but is not shown in the profile)
- in case that only one (*e.g.* mean) control group risk is chosen to be displayed, information about the range of control-group risks in individual studies may be used as a reference available for updating the evidence profile (it is less important when creating Summary of Findings tables for systematic reviews, because this information is available in the review itself)

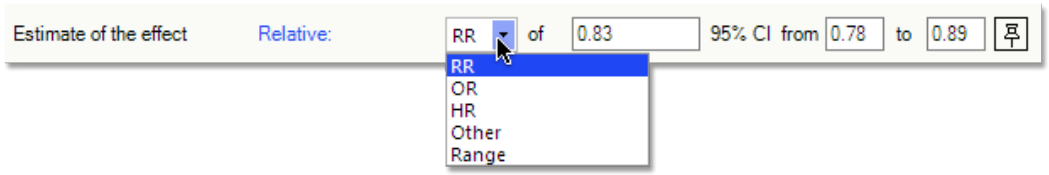
5.5.1.5 Relative effect

About Relative Effect

The relative effect for a dichotomous outcome from a single study or a meta-analysis will typically be a risk ratio (relative risk), odds ratio, or occasionally a hazard ratio.

You may want to present a relative effect measure you found in the literature you use to develop a GRADE evidence profile or to convert different relative effect measures). From the drop down menu you can select the relative effect used in the meta-analysis or reliteve effects of one or more studies if no pooled estimate is available. There are 5 options:

- RR (Risk Ratio)
- OR (Odds Ratio)
- HR (Hazard Ratio)
- Range (when meta-analysis was not conducted and a range of relative effects from studies can be presented; you should specify the type of effect measure that was used, *e.g.* RR or OR)
- Other

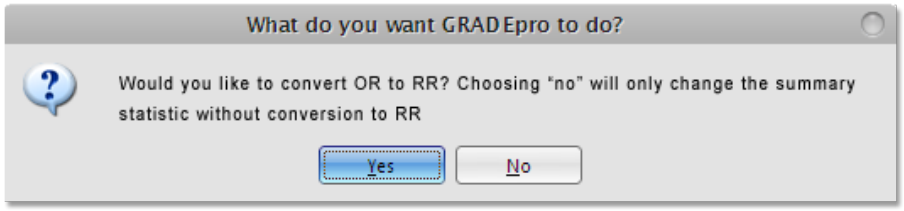


You can enter the point estimate and the confidence limits (a range of point estimates if there were several studies but no meta-analysis).

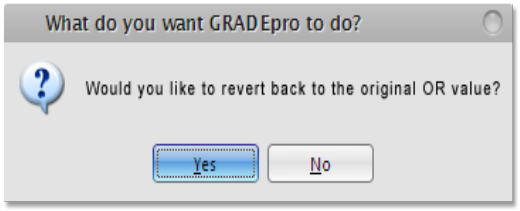
You should add footnotes to provide more information about the estimate of the effect.

When results are not pooled and a range of effects cannot be presented, you can briefly summarize the qualitative assessment of the pattern of the results in a footnote. However, you should avoid "vote counts" — reporting numbers of "positive" and "negative" studies, because they are not informative.

Note
GRADEpro can **convert an Odds Ratio or Hazard Ratio** (which was previously entered or imported into GRADEpro) **into an RR**. After entering the values for OR or HR select RR from the drop down menu. A pop-up window will include a question whether you would like to convert OR or HR to RR. If you want to convert choose **YES**, if you want just to change the label for summary statistics choose **NO**.



The original OR or HR is saved in GRADEpro so that after converting them to RR, you can revert back to an original OR or HR. Select OR or HR in the drop down menu and choose YES to the pop-up window "Would you like to revert back the original value?"



The formulae GRADEpro uses to convert OR or HR to RR.

5.5.1.6 Absolute effect

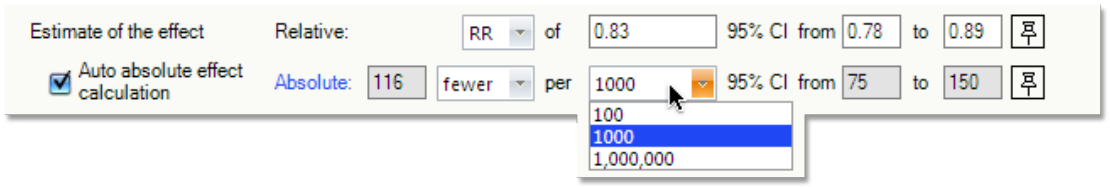
About the Absolute Effect

The **absolute measure of intervention effects** is a difference between the baseline risk of an outcome (*e.g.* in patients receiving control intervention or estimated in the observational studies) and the risk of outcome after the intervention is applied, *i.e.* the risk of an outcome in people who were exposed or received an intervention. Absolute effect is based on the relative magnitude of an effect and baseline risk.

In GRADEpro **Auto absolute effect calculation** option is selected by default. GRADEpro will **automatically calculate** the absolute effect based on the baseline risk and the relative effect size. GRADEpro will calculate the absolute effect risk once at least one of the baseline risk values is provided and the magnitude of the relative effect (RR, OR or HR) has been entered. Absolute effect values are automatically entered into the GRAEDE evidence profile.

You can switch **Auto absolute effect calculation** option off and provide your own estimate of absolute effect, e.g. when a risk difference meta-analysis was done.

From the drop down menu you should choose a **denominator to present risks**. You can present baseline risk, risk after the intervention or exposure, and the calculated absolute effect as number of subjects per 100, per 1000, or per 1,000,000. By default, risk is presented **per 1000** people. You can choose to present risk per 1,000,000 if the events are rare or per 100 if the events are more frequent.



Note

Auto absolute effect calculation option and **calculation of absolute effect** **do not pertain to SoF tables** and can be ignored. This function is used for creating GRADE evidence profiles.

5.5.1.6.1 Calculation of absolute effect

About Calculation of Absolute Effect in GRADEpro

GRADEpro will automatically calculate absolute measure of intervention effects in GRADE evidence profile based on the illustrative baseline risk and the relative effect size according to the following formulae.

Absolute effect from Risk Ratio (RR)

Absolute effect (per 1000 people) = 1000 × | CER × (1 — RR)|
where
CER — Control Event Rate (CER = number of people with event in control groups ÷ total number of people in control groups)
RR — Risk Ratio

Example

Risk ratio (RR) from meta-analysis is 0.83 (95% CI, 0.78 to 0.89),
The control event rate (CER) is 40% (400 per 1000 or 0.4).
therefore
Absolute effect = 1000 × 0.4 × (1 — 0.83) = 68 fewer per 1000.

Absolute effect from Odds Ratio (OR)

The odds ratio (OR) is first converted to risk ratio (RR) and the corresponding risk is then calculated as above.
The formula for the conversion of odds ratio to risk ratio uses control event rate (CER) from the meta-analysis:

$$RR = \frac{OR}{1 - CER \times (1 - OR)}$$

where
CER — Control Event Rate (CER = number of people with event in control groups ÷ total number of people in control groups)
OR — odds ratio
RR — Risk Ratio

Example

Odds ratio (OR) from meta-analysis is 0.64 (95% CI, 0.47 to 0.89),
CER = 112 ÷ 438 = 25.6% or 0.256
therefore
RR = 0.64 ÷ (1—0.256 × (1—0.64)) = 0.70
therefore
Absolute effect = 1000 × 0.256 × (1 — 0.70) = 77 fewer per 1000.

Absolute effect from Hazard Ratio (HR)

The hazard ratio (HR) is first converted to risk ratio (RR) and the corresponding risk is then calculated as above from RR.
The formula for the conversion of hazard ratio to risk ratio uses control event rate (CER) from the meta-analysis:

$$RR = \frac{1 - e^{HR \times \ln(1 - CER)}}{CER}$$

where
CER — Control Event Rate (CER = number of people with event in control groups ÷ total number of people in control groups)
HR — hazard ratio
RR — Risk Ratio

5.5.2 Continuous outcomes

Completing Information about Continuous Outcomes

The term "continuous" in statistics conventionally refers to data that can take any value in a specified range (Cochrane Handbook Section 9.2.3). Examples of truly continuous data are distance, area and volume. In practice, one can use the same statistical methods for other types of data, most commonly measurement scales and counts of large numbers of events (Cochrane Handbook Section 9.2.4).
The mean difference (MD) or standardised mean difference (SMD) are the two key pooled estimates of effect for continuous outcomes measured on a continuous or ordinal scale. The presentation of the results for these outcomes depends on the type of pooled estimate calculated.

Note
When completing the information for continuous outcomes, it is important to consider the differences in scores that you will present in the GRADE evidence profile. You can present the difference in scores between the control group/assumed risk and the intervention/corresponding risk group for

- change in scores from baseline to end of study (change scores)
- scores from the end of the study (final values).

When naming the outcome indicate if it is considered an absolute value or a change score.
For example, specify **Pain** versus **Change in Pain** or **Quality of Life** versus **Change in Quality of Life**.

Other informative ways of presenting continuous outcomes are the **ratio of means** (*e.g.* the ratio of the mean in weight gain in one group compared to another). A special case that can be considered a continuous outcome is a **comparison of rates or rate ratios** (*e.g.* the number of disease exacerbations per patient or the number of new polyps per patient in one group compared to another).

You will need to enter the following information for a continuous outcome:

- length of follow-up
- number of participants
- assumed risk (final values or change score in control group)
- estimate of the effect
- measurement scale

5.5.2.1 Length of Follow-up

Specifying the Length of Follow-up

You can provide information on the length of the follow-up in the studies that reported each outcome. This is important, since the interpretation of the observed effect depends on the time in which the events occurred.

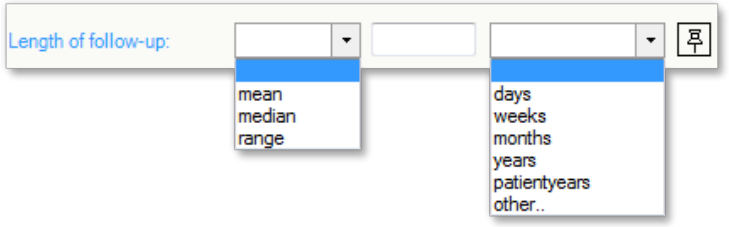
You will have to make **judgments across studies**, because it is unlikely that the duration of follow-up was the same in all studies.

Note
The **duration of intervention** (which should be described in the intervention type, *e.g.* treatment with glucosamine for 8 weeks) may differ from the **duration of follow-up** (*e.g.* assessment of health related quality of life or function after 1 year).

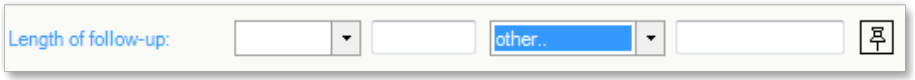
You should provide the information on the length of follow-up for each outcome separately since they may have been measured in different periods.

To provide the information on the time of follow-up:

- Choose **mean**, **median**, or **range** from a drop-down list
- Type a **number** (for mean or median) or a range (e.g. **4 to 7**) in the box
- Choose the unit of time from a drop-down list (**days**, **weeks**, **months**, **years**, **patient-years**, or **other...**)



If you choose other... measure of time an additional box will appear allowing you to specify the unit (*e.g.* hours, etc.)



For authors of systematic reviews

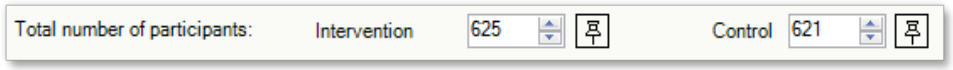
The length of follow-up at which time the outcome was measured will always need to be entered regardless of whether data were imported from a Review Manager 5.0 file or not.

5.5.2.2 Participants

Providing Number of Participants

You will need to enter a total number of participants in control groups and intervention/treatment groups in the studies that contribute to the results. Usually this information can be obtained directly from a meta-analysis graph (forest plot).

For a **continuous outcome** you should specify **total number of participants** in the intervention/treatment and control groups.



For a **dichotomous outcome** you should also specify the **number of participants (patients) in whom the event occurred**.

In the **grey boxes to the right** you will see the **percentage** of total subjects in whom the outcome occurred. The check-box next to the control group event rate lets you choose this rate as a basis for calculation of the absolute effect observed in the studies.

> *For authors of systematic reviews*
>
> When **importing** from Review Manager 5.0 file this information will be **automatically entered** into GRADEpro.

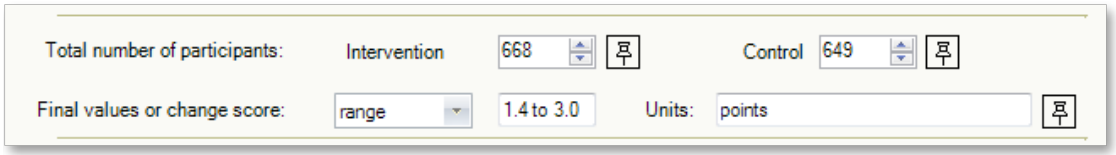
5.5.2.3 Control score

About Assumed Risk (Score)

The Assumed Risk (Score) is the score of the participants who received the control intervention (in the context of systematic review or single study) or for whom the recommendation is intended (in the context of clinical practice guidelines).

Unlike dichotomous outcomes that are measured once at the end of a study, continuous variables are measured both at the beginning (baseline – before observation started or interventions were administered) and at the end of a study. Therefore, continuous outcomes can be expressed either as a **change in scores from baseline to end of a study** (*i.e.* change score) or as **final measurements** at the end of the study (*i.e.* final values). **The Assumed Risk (Score) is either a final value or change score in the control group.** Choice of the presentation of control score in the evidence table will depend on the type of data used in the meta-analysis. You should specify in the outcome name or in a footnote which of the two presentations was used (*e.g.* pain intensity *versus* change in pain intensity).

You can estimate the Assumed Risk (Score) by assessing typical scores in different patient groups or at different lengths of follow-up. Ideally, these groups would reflect patients that clinicians can easily identify on the basis of their presenting features. You can derive the scores in the controls either directly from a systematic review from which you obtained the estimate of the intervention effect or you can extract it from observational studies of patients similar to those for whom the intervention is intended.



---

1. Assumed risk (score) when an effect was expressed as a MEAN DIFFERENCE (MD)

You can provide a range, median or mean of the final values or change scores in the control groups.

- range — it can be the highest and the lowest estimate of the scores in the control groups
  Note: **if there are "outliers"** another approach for obtaining high and low estimates would be to use the **second lowest and second highest** control score from the studies.
- median — the mid level score can be based on the median of the scores in the control groups across studies in the systematic review or on data from representative observational studies
- mean — the mid level risk can be based on the weighted mean of the scores in the control groups across studies in the systematic review or on data from representative observational studies.

You should provide a footnote clarifying the source of the assumed risk (score) used.

You should also provide the **units of the scale on which the scores were measured** (*e.g.* kg, points, days, etc.). These units will appear in the evidence profile (SoF table).

Note
Final values or change scores will **not** be automatically imported into GRADEpro from Review Manager 5 file.

---

2. Assumed risk (score) when an effect was expressed as a STANDARDIZED MEAN DIFFERENCE (SMD)

There are three alternative methods for presenting a SMD (» Re-expressing SMD) and therefore three different presentations of final values or change scores in the control groups.

1. when using rules of thumb for effect sizes — **you do not fill in information about the final values or change scores**, you present only the SMD with the corresponding confidence interval
2. when transforming to odds ratio — if you "dichotomize" results of continuous outcome by transforming SMD to OR (or RR), you can apply these numbers to an assumed risk similar to the calculations for dichotomous outcomes. You will need to change the type of the outcome from continuous to dichotomous on the Quality Assessment editing pane and then enter the data according to presentations for dichotomous outcomes. If you choose this option, you can also provide up to 3 risk estimates for different risk groups
3. when using a familiar instrument — for the "back translation" of a SMD to a familiar instrument you can present a mean, median, or range of final values or change scores similar to the presentation for MD (see above). The mean, range or median, however, will be from the studies that used the same familiar instrument.

Note
All of these alternative methods for presenting a SMD have limitations since they require several statistical assumptions that do not hold under all circumstances. However, these approximations facilitate interpretable presentations of results.

---

3. Assumed risk (score) when an effect was expressed as a RATIO OF MEANS

You do not fill in information about the final values or change scores in control groups.

5.5.2.4 Estimate of the effect

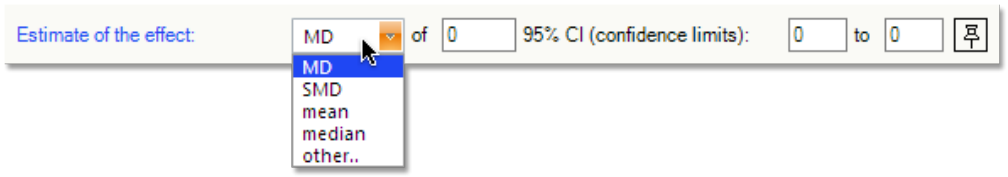
About Estimate of the Effect

The estimate of the effect can be presented in a variety of ways. There are 5 options:

- mean difference (MD)
- standardised mean difference (SMD)
- mean
- median

- other

You can select the effect used in the meta-analysis from the drop down menu.



---

1. When using a MD, mean, or median

You can enter the point estimate and the confidence intervals. You can add footnote to provide more information about the effect and confidence intervals as necessary.

> *For authors of systematic reviews*
>
> If you imported data from a Review Manager 5 file the estimate and confidence intervals will be automatically imported.

---

2. When re-expressing SMDs

You will need to enter or edit the data depending on how the SMD is re-expressed.

1. when using rules of thumb for effect sizes — you should select MD from the drop down menu and fill in the **MD** and confidence intervals. This information is presented in the corresponding risk column of the outcome. See below for suggestions for comments. A footnote should be provided.
2. when transforming to odds ratio — you should calculate OR from the SMD and then apply the OR as with dichotomous outcomes. You will need to change the type of outcome to **dichotomous** on the Quality Assessment editing pane and then enter the data according to presentations for dichotomous outcomes. You should include the original SMD value in the comments section (*e.g.* "Numbers estimated using a standardised mean difference of XX (95% CI YY to ZZ)")
3. when using a familiar instrument — for the "back translation" of an SMD to a familiar instrument you should choose a representative study and then calculate the difference by multiplying the SMD by the SD (standard deviation) of the mean change in the control group in this study. You should select **MD** as the type of estimate and then enter the difference and the confidence intervals. These numbers will be displayed in the assumed risk and corresponding risk columns similar to the MD presentation.

   Note

   If you import data from a Review Manager 5 file you will need to select **MD** in the drop down menu and then change the numbers that were originally imported as SMD into the difference and confidence intervals calculated using data from representative study.

---

3. When using the ratio of means

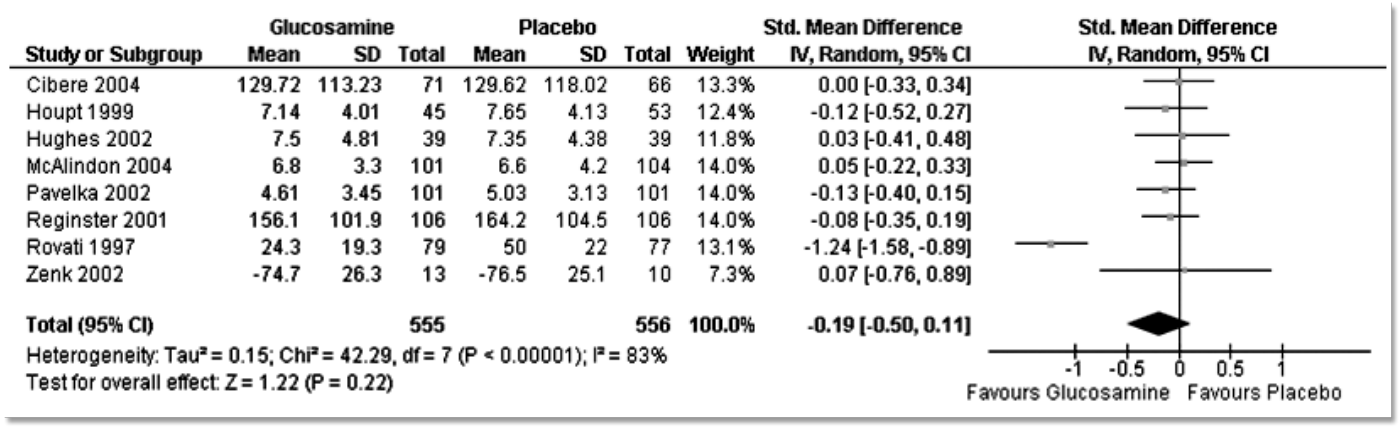
The ratio of means represents the weighted average of the mean scores in the intervention group divided by mean score in comparison group. Only the relative effect is given and the final values or change score in control groups or in the intervention group are not given. You should choose **other...** from a drop-down menu and enter information about the ratio of means directly into the comments column.

5.5.2.4.1 Re-expressing SMD

About Re-expressing SMD

When studies have used **different instruments** to measure the same construct, a standardized difference in means (SMD) may be used in meta-analysis for combining continuous data. The SMD expresses the intervention effect in **standard deviation (SD) units** rather than the original units of measurement. The mean difference (whether change from baseline to end of study, or end of study value) is standardized by dividing it by the standard deviation in the control group in this study. The standardized means from the individual studies are then combined in meta-analysis to calculate SMD. Consequently, the value of SMD depends on both the **size of the effect** (the difference between means) and the standard deviation of the outcomes (the inherent **variability among participants**).

Forest plot used in examples below.



There are three options for re-expressing the SMD facilitating its interpretability:

---

1. Re-expressing SMDs using rules of thumb for effect sizes

Rules of thumb exist for interpreting SMDs or "effect sizes". If you choose this mode of presenting SMD **you should include a rule of thumb** in a footnote. You should bear in mind that some methodologists believe that such interpretations are problematic, because patient importance of a finding is context-dependent and not amenable to generic statements.

Rule of thumb according to Cohen's interpretation of effect size

- 0.2 represents a small effect
- 0.5 represents a moderate effect
- 0.8 represents a large effect

There are variations of Cohen's interpretation. An example might be:

- <0.41 represents a small effect
- 0.40 to 0.70 represents a moderate effect
- >0.70 represents a large effect.

A sample GRADE evidence profile presenting SMD from the above example using Cohen's interpretation of effect size

Quality assessment							Summary of findings					
							No of patients		Effect			
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	glucosamine	no treatment	Relative (95% CI)	Absolute	Quality	Importance
Pain (follow-up mean 3 months; measured with: Measured with different scales in the different studies. Lower scores mean less pain.; Better indicated by less)												
8	randomised trial	serious	serious[1]	no serious indirectness	serious[2]	none	555	556	-	MD -0.19 (-0.5 to 0.11)[3]	⊕⊕◯◯ LOW	CRITICAL

[1] High heterogeneity unexplained.

[2]

The confidence interval does not rule out a null effect or harm.

[3] This is a difference in standard deviations. A standard deviation of 0.2 represents a small difference between groups.

---

2. Re-expressing SMDs by transformation to odds ratio

A transformation of a SMD to an odds ratio (OR) is possible. Due to the underlying assumptions to make this conversion, the results are only an approximation. To calculate OR use the formula:

$$\ln(OR) = \frac{\Pi}{\sqrt{3}} SMD$$

where Π/√3 is approximately 1.8138

The estimated odds ratio can then be entered similarly as for a dichotomous outcome. The assumed risk (control group risk) refers to the proportion of people who have improved by some unspecified amount (or those without an event) in the continuous outcome ("responders"). GRADEprofiler can then automatically calculate the corresponding risk based on the assumed risk entered and present the results as dichotomous outcome. You should add a comment such as, "numbers estimated using a standardised mean difference of XX (95% CI YY to ZZ)"). If you select this option you will be able to choose more than one assumed risk value as for other dichotomous outcomes.

In the above example the SMD was 0.19 which multiplied by 1.8138 gives 0.34. If ln (OR) = 0.34 then OR = 1.41. The assumed risk was 0.9.

Note

In the SoF below the outcome is number of people who had little or no pain (NOT number of people with pain).

A sample GRADE evidence profile presenting SMD from the above example using transformation to odds ratio

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	glucosamine	no treatment	Relative (95% CI)	Absolute		
Little or no pain (follow-up mean 3 months; measured with: different scales in the different studies)												
8	randomised trial	serious	serious[1]	no serious indirectness	serious[2]	none	0/555[3]	9%	OR 1.41 (1.22 to 2.48)[4]	3 more per 100	⊕⊕◯◯ LOW	CRITICAL

[1] High heterogeneity unexplained.

[2] The confidence interval does not rule out a null effect or harm.

[3] There were 555 patients in glucosamine groups and 556 in no treatment groups. Effect was estimated using a standardised mean difference.

[4] Numbers estimated using a standardised mean difference of -0.19 (-0.50 to 0.11)

---

3. Re-expressing SMDs using a familiar instrument

Because the SMD is based on standardized means from the included studies and not a specific scale, it is unit-less. This makes the interpretation of the effect very difficult. To better understand the effect it can be re-expressed by applying the calculated SMD back into one of the original studies and depicted on the scale used in that study. To back transform the SMD to familiar scale

- select a study included in the original meta-analysis that is representative of the population and intervention and at a low risk of bias
- multiply the standard deviation of the control group (end of study mean or mean change from baseline to end of study) by the pooled SMD

This resulting number is the mean difference (MD) and can be presented in the Summary of Findings table in the format of MD (mean difference) for the scale used in the representative study.

Note

One should interpret such results with caution since back-translation of the effect size is based on the results of only 1 study.

A sample GRADE evidence profile presenting SMD from the above example using back-translation to a familiar instrument

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of						Other		no	Relative			

studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	considerations	glucosamine	treatment	(95% CI)	Absolute		
Pain (follow-up mean 3 months; measured with: WOMAC[1,2]; range of scores: 0-20; Better indicated by less)												
8	randomised trial	serious	serious[3]	no serious indirectness	serious[4]	none	555	556	-	MD -0.8 (-2.1 to 0.5)[2]	⊕⊕◯◯ LOW	CRITICAL

[1] Four of the eight studies used the WOMAC pain subscale.

[2] Scores estimated using a standardised mean difference of -0.19 (-0.50 to 0.11)

[3] High heterogeneity unexplained.

[4] The confidence interval does not rule out a null effect or harm.

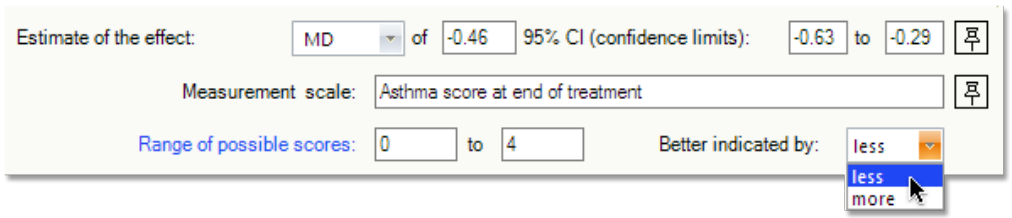
5.5.2.5 Measurement scale

Describing Measurement Scale

Measurement scale

You can enter

- specific name of the measurement scale or the familiar instrument used to re-express SMD (*e.g.* SF-36, WOMAC, etc.)
- description of how the outcome was measured (*e.g.* visual analogue scale)
- statement that the "outcome was measured on different scales in different studies" in case of SMD re-expressed as an effect size



Range of possible scores

You should enter the range of possible scores on the scale used to measure the outcome. You can provide a range of numbers or numbers and text to describe the scores. For example, 0 to 100 or "0 (no difficulty) to 5 (severe difficulty)"

Better indicated by

You can fill in the direction of the effect on the scale. However, this information does not appear in the Summary of Findings table.

For authors of systematic reviews

When importing from a Review Manager 5 file any of the above information is **not imported**.

5.6 GRADE evidence profiles for observational studies

GRADE evidence profiles for observational studies

GRADE evidence profiles for observational studies **may require additional editing** which can be done in word processor. This situation can occur when outcomes are derived from observational studies that include **case-control studies** as the presentation of results can differ.

Examples

1. Mixed presentation of cohort and case control studies

Question: Should rofecoxib be used for arthritis mixed observational studies?
Settings: primary care
Bibliography: McGettigan & Henry JAMA 2006; 296:1633-1644

Quality assessment							Summary of findings				Importance
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	rofecoxib	control	Relative (95% CI)	Absolute	
Acute myocardial infarction in case control studies (exposure mean 5 years[1]; assessed with: administrative databases)											
12	observational study[2]	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	dose response gradient[3]	Xy cases Xy controls (cohort studies: 1100 events /755203 exposed and 6000/5947203 events/non exposed)[4]		RR 1.35 (1.15 to 1.59)	3 more per 1,000	⊕⊕⊕○ MODERATE	CRITICAL

[1] exposure and follow up differed widely in these studies – not actual data. The average duration of exposure was xy years in the cohort studies with the same duration of follow-up and exposure was xy years in the case-control studies, respectively

[2] 10 case-control studies and 2 cohort studies are included

[3] A dose response gradient was present

[4] only for example - actual numbers not given in review

2. Presentation of case control studies (in this example the cells referring to the population in the outcome row are merged)

Question: Should rofecoxib be used for arthritis in case control studies?
Settings: primary care
Bibliography: McGettigan & Henry JAMA 2006; 296:1633-1644

Quality assessment							Summary of findings				Importance
							No of patients		Effect		Quality
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	rofecoxib	control	Relative (95% CI)	Absolute	
Acute myocardial infarction in case control studies (exposure mean 5 years[1]; assessed with: administrative databases)											
10[2]	observational study[2]	no serious limitations	no serious inconsistency	no serious indirectness	no serious imprecision	dose response gradient[3]	Xy cases Xy controls[4]		RR 1.31 (1.31 to 1.46)	3 more per 1,000	⊕⊕⊕○ MODERATE	CRITICAL

[1] average duration of exposure – not actual data
[2] 10 case-control studies are included
[3] a dose response gradient was present
[4] only for example - actual numbers not given in review

6. Creating Summary of Findings Tables

Creating Summary of Findings Tables

Authors of Cochrane systematic reviews can use GRADEpro to create a Summary of Findings (SoF) table for their systematic review. Summary of Findings tables present the main findings of a systematic review in a simple table format.

Usually, authors of Cochrane reviews will import the data from their RevMan file. Other users of GRADEpro will have to prepare full GRADE evidence profiles.

There are 3 key parts of a SoF table:

- information about the review
- summary of the statistical results
- grade of the quality of evidence.

Authors of systematic reviews can use GRADEpro to create a SoF table in two ways:

- import most of the data from a Cochrane systematic review from Review Manager (RevMan) 5.0 file
- enter data manually.

Regardless of how you enter the data, you will need to edit data and manually enter your assessment of the quality of evidence for each outcome.
When SoF table is finished, ypu can save the GRADEpro file for later use, export the SoF table and/or import it into RevMan.

Overall tasks to create a SoF table

1. Open GRADEpro.
2. Choose **New Profile** in the welcome screen and name the file.
3. Create profile group
   a. if you are working with a Review Manager file, import the data from RevMan file. From the **File** menu choose **Import From** » **Review Manager** and select the Review Manager 5.0 file from which you want to import the data.
   b. if you are manually entering data from a systematic review, name the profile group (usually the overall topic of the review).
4. Create a profile (comparison)
   a. if you imported data from Review Manager file, your 'comparisons' from the review will be translated into health care questions defining the profiles
   b. if you enter data manually create a profile for each comparison (detailed health care question) in the review
5. Add or edit the information about the comparison (health care question and information about bibliography and authors).
6. Create an outcome
   a. if you imported data from Review Manager file, the outcomes from the meta-analysis will be imported, however you will still have to manually add and manage outcomes that you consider important to patients, but were not combined in the review
   b. if you enter data manually create an outcome
7. Add, move or delete outcomes under each profile (comparison).
8. Select an outcome. For each outcome there are 2 sections: Summary of Findings screen and the Quality Assessment screen.
9. Select the Quality Assessment screen. Complete it by first confirming number and type of studies, then assess the quality of evidence for the outcome. Downgrade or upgrade evidence

according to GRADE criteria and enter footnotes when necessary or required.

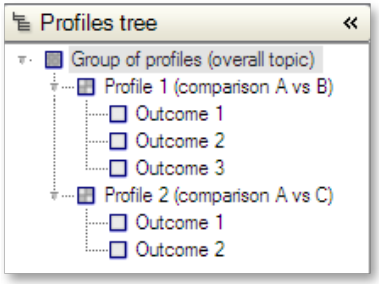
10. Select the Summary of Findings screen. Confirm the data that were imported or add data manually. If data were imported, enter optional data (e.g. assumed risks, scale descriptions; follow-up; event rate denominator, and footnotes).
11. Repeat (#9 and #10) for all outcomes in the profile.
12. If, there is more than 1 comparison of importance in the review, repeat (#4 to #11) for all profiles to create another SoF table.
13. Preview Summary of Findings table, double check presentation and edit if necessary. **Do not skip this step.**
14. Export Summary of Findings table. Choose the profile (comparison) to present in the table. A maximum of 7 patient-important outcomes should be presented in the SoF table.
    a. you can export SoF in Review Manager format that may be later imported into Review Manager 5
    b. you can also export to another type of file (Microsoft® Office Word, HTML, picture, etc).

6.1 Creating group of profiles

Creating a Group of Profiles

Groups of profiles represent an **overall topic** of a systematic review or overall topic of a health care recommendation, that may involve consideration of several health care questions (comparisons presented in evidence profiles) or an overview of reviews.

Any GRADEprofiler file has to contain at least one group of profiles at the root of the tree of profiles.



The **group of profiles** allows organizing several evidence profiles regarding a similar topic in one batch. It facilitates viewing, organizing, and saving evidence profiles for similar health care questions. It also allows exporting Overview of Reviews table.

You may add an additional group of profiles within the same GRADEprofiler file whenever you want to put together evidence tables containing information on another health care problem.

To create a Group of Profiles:

- Choose **Profile Group** from the **Add** menu or click on a toolbar.
- Type the name of a Profile Group.
- Click **Add profile**.

A new **Add Profile** dialog box will automatically appear.

Example

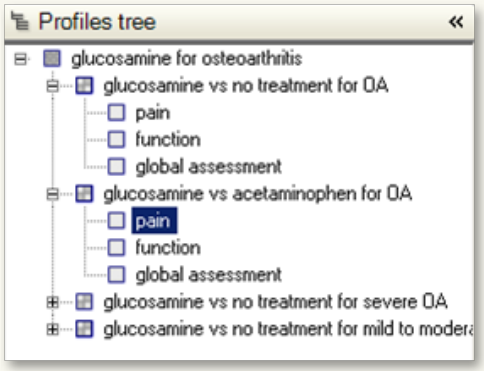
You are interested in creating several evidence profiles concerning different management options for lung cancer. One Group of Profiles could contain health care questions about surgical methods and the other about medical management of lung cancer.

For authors of systematic reviews

In GRADEpro, a systematic review will usually be organised into three levels:

- **Profile Group**: the main topic of the review. For example, glucosamine and other interventions for osteoarthritis.
- **Profiles**: the (multiple) comparisons in the review. For example, glucosamine versus no treatment, and glucosamine versus acetaminophen OR glucosamine versus no treatment in severe osteoarthritis and glucosamine versus no treatment in mild to moderate osteoarthritis. Since a review may have more than one comparison and also comparisons with subgroup analyses, there will often be more than one profile in one review. Each profile is set up as a more specific or focused health care question than the Profile Group.
- **Outcomes**: the outcomes measured in each comparison. For example, pain, function, global assessment, and adverse effects.

The tree for the example above would appear in GRADEpro as



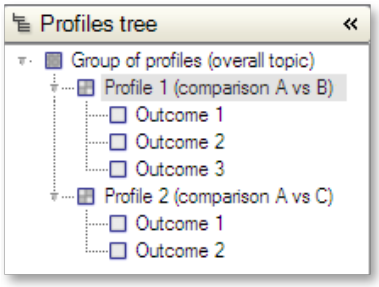
For authors that import data from RevMan the profile group, profiles, and outcomes will be imported. It may be necessary to move, delete or add other profile groups, profiles, and outcomes depending on how the comparisons and outcomes were created and organised originally in RevMan.

6.2 Creating a profile

Creating a Profile

Evidence profile (GRADE evidence profile or Cochrane Summary of Findings table) is a table summarizing the information about the outcomes for a focused health care question. It provides simple and transparent summary of available evidence about a particular health care management option.

A profile is a subdivision of a profile group . An evidence profile is defined by a specific health care question following the PICO format. It **represents a single comparison** (*e.g.* intervention versus placebo) for which multiple outcomes may have been measured.



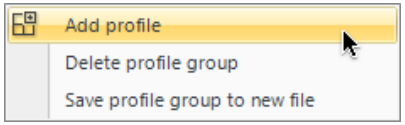
To create a Profile:

- Choose **Profile** from **Add** menu.
- Specify a health care question.
- Type or paste bibliographic information.
- Type the name of the author of the profile.

Note

There are four possibilities to create a new Profile:

- Choose **Profile** from the **Add** menu
- Click on a toolbar
- Press **Ctrl + Shift + P**
- Right-click on a Group of Profiles or on a Profile in the tree on the left-hand side and choose **Add Profile**



6.3 Creating and managing outcomes

Creating and Managing Outcomes

You can specify one or more outcomes for any health care question indicated in the profile.

In most clinical situations patients are interested in many outcomes (*e.g.* patients with asthma may be interested in preventing an exacerbation requiring hospitalization, need for administration of oral glucocorticosteroid, aggravation of symptoms disturbing their sleep, etc.).

It is important to include in the evidence table (Summary of Findings table or GRADE evidence profile) **all outcomes that are ultimately important to patients** regardless if they were measured in the studies or not. Ideally authors of a systematic review would **specify all patient-important outcomes before conducting a review** and those making recommendations should specify these outcomes at the beginning of guideline development process. Outcomes that were either not reported, not measured, or not pooled, but considered important to patients and would affect decision making, should still be included in the evidence table (SoF table), although one can neither grade quality of supporting evidence (because it is not

available) nor provide the estimated magnitude of effect. For these outcomes the ratings are left blank (see examples below).

SoF table

A compared to B for some health problem

Patient or population: patients with some health problem
Settings:
Intervention: A
Comparison: B

Outcomes	Illustrative comparative risks* (95% CI) Assumed risk B	Corresponding risk A	Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
Quality of life - not measured	See comment	See comment	Not estimable	-	See comment	Quality of life was not measured it any of the studies

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval;

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

GRADE evidence profile

Question: Should A vs B be used for some health problem?
Settings:
Bibliography:

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	A	B	Relative (95% CI)	Absolute		
Quality of life - not measured												
0	-	-	-	-	-	none	0	0	-	-		CRITICAL

Information about the outcome

Each outcome in the evidence table (SoF table) consists of two sets of data: quality of the evidence and the results of the studies. There are two panes in GRADEpro to enter and edit that data:

1. quality assessment of the evidence
2. summary of findings or the results

You will need to add, edit, or delete information in each edit pane to fill in the evidence (SoF) table.

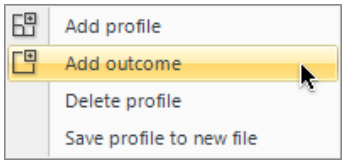
To create an Outcome:

- Choose **Outcome** from **Add** menu.
- Name the outcome.
- Choose the type of the outcome (dichotomous or continuous).
-

Rate importance of the outcome.

There are four possibilities to create a new Outcome:

- Choose **Outcome** from **Add** menu
- Click on a toolbar
- Press **Ctrl + Shift + O**
- Right-click on a Profile in the tree on the left-hand side and choose **Add Outcome**



For authors of systematic reviews

You will likely not include all of the outcomes from the review in the Summary of Findings table (a **maximum of 7 outcomes are recommended**).

When importing data from a Review Manager 5.0 file

1. All outcomes that are summarized in with a **meta-analysis** (including subgroups) will be imported into GRADEpro (see below for exceptions).
2. Outcomes with **sub-totals only** will not be imported. However, the subgroup outcomes if totalled will be imported.
3. Outcomes with no totals are imported. Authors may want to choose 'single study' when defining the status of the effect measure of an outcome.
4. Authors will need to move, add or delete outcomes depending on which outcomes were imported and which outcomes are important to patients. You can order the outcomes by their importance by draging them in the tree pane.
5. Imported from the Review Manager 5.0 file are: information about the type of outcome, number of studies and some of the data from the meta-analysis such as control and intervention events, point estimates. Other data will need to be entered (see subsections under Completing Quality Assessment for an outcome and Completing the Summary of Findings for an outcome).

6.4 Completing the quality assessment for an outcome

Completing Quality Assessment for an Outcome

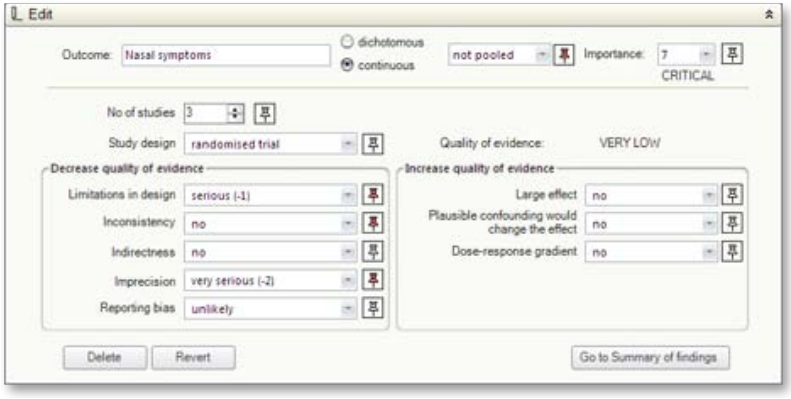
You should **assess the quality of the evidence for each outcome**, with the **exception** of outcomes for which there are no studies that measured it (**not measured**) or no studies reported this outcome (**not reported**).

You should carefully consider and rate all the factors that may influence the quality of evidence. While doing so, you should bear in mind that down- and upgrading for specific quality factors should be done in the context of all of the factors that influence the quality of evidence. Having downgraded for one quality criterion may influence how the next quality criterion is dealt with. Thus, it could influence the threshold for downgrading of another criterion. While it is acceptable to avoid judgments that might be overly severe, the decisions should be made transparent. For example, borderline situation in which it would be reasonable either to downgrade for study limitations, or not to do so may exist. This illustrates that the great merit of GRADE is not that it ensures consistency of conclusions, but that it requires explicit and transparent judgments. In any case you should explain your decision in a footnote.

The editing pane is divided into:

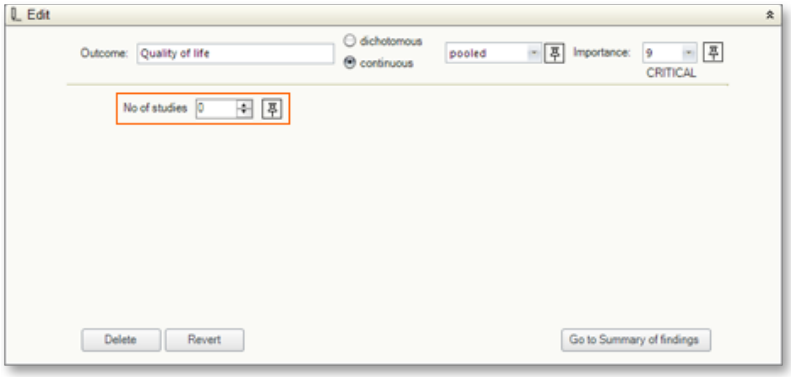
- information about the outcome
- quality assessment
  - number of studies

- type of studies
- factors that affect the quality of the evidence

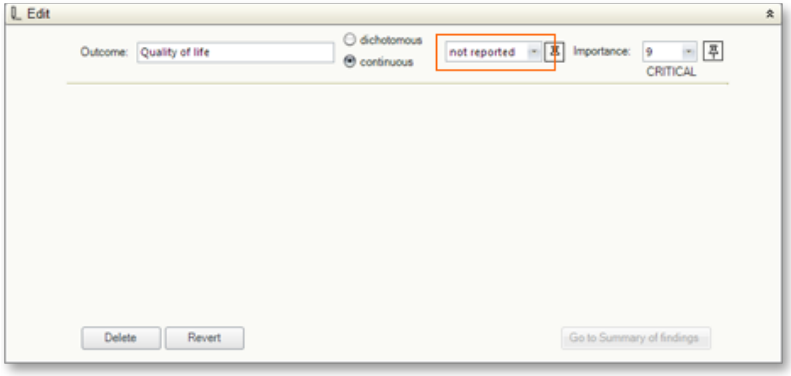


Note

The editing pane for Quality Assessment will appear **empty until you choose at least one study and type of studies** contributing to the results.



It will also be empty when the outcome was either not measured or not reported. In this case you will not be able to proceed to Summary of findings pane since there are no findings to report.



Factors that influence Quality of Evidence in Sof Tables

Decrease	Increase
Study limitations	Large magnitude of effect

Inconsistency of results	All plausible confounding would reduce the demonstrated effect
Indirectness of evidence	Dose-response gradient
Imprecision	
Publication bias	

Related topics

- Detailed description of factors influencing quality of evidence
- Diagram for grading the quality of evidence using the GRADE approach

6.4.1 Reporting the number of studies

Reporting the Number of Studies

You should specify the total number of studies included in the systematic review that examined this outcome.

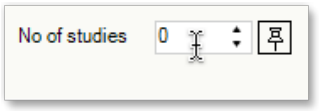
Specifying the number of studies informs users whether the outcome was addressed in:

- no study — there are no published data that fulfilled the inclusion criteria for the question asked
- only in one study
- many studies.

You should specify the number of studies everytime you add a new outcome.

To specify the number of studies:

- Type a natural number or use the up and down arrows in a text box
- Explain your choice in a footnote if you think it may not be clear why you have chosen that number.



For authors of systematic reviews

When importing data from a Review Manager 5.0 file, the number of studies that were included in the meta-analysis for the given comparison for a specific outcome will be automatically entered here.

6.4.2 Choosing study design

Choosing study design

You can specify the basic design of studies included in the systematic review that examined this outcome. Choose a basic study design that is an essential determinant of the quality of evidence:

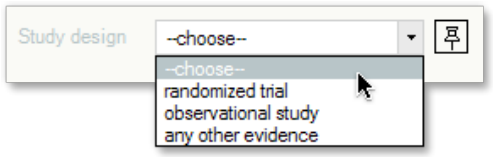
- randomised trial
- observational study
- any other evidence (case series or case reports)

Note
The category referring to **any other evidence** such as case series or case reports is currently included, but users may choose to include this category in observational studies and downgrade in the detailed assessment of study design and limitations for the absence of an independent control group.

Choosing the basic study design is a prerequisite to grading the quality of evidence for any outcome.

To choose the basic study design:

- Choose a study design from a drop-down list
- Add any other information in a footnote if you think it may be important for others to know this.



6.4.3 Identifying type of outcome

About Types of Outcomes

There are two types of outcomes: **dichotomous** and **continuous**, that require different presentation of findings, because their summary statistics differ.

Dichotomous (binary) outcomes

Each individual's outcome may have one of only two possible categorical responses (e.g. dead or alive, myocardial infarction or no myocardial infarction, etc.).

Continuous outcomes

The term 'continuous' in statistics conventionally refers to data that can take any value in a specified range. A continuous outcome may be either:

- "truly" **continuous** (e.g. blood glucose concentration or weight)
- **ordinal** (e.g. score on a Hamilton depression scale or Chronic Respiratory Questionnaire)

Summary statistic and the presentation of results for both types of "continuous" outcomes are usually the same.

> *For authors of systematic reviews*
>
> The type of an outcome will determine the data that you will need to enter in GRADEpro. For authors who import data from a Review Manager 5.0 file, the type of outcome will be determined automatically.

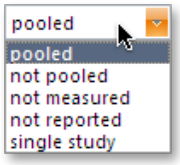
6.4.4 Specifying status of the analysis for an outcome

Specifying the Status of the Analysis for an Outcome

Status of the analysis for an outcome refers to one of a number of possibilities **how an effect measure is be presented**.

You can specify the status of a particular analysis for an outcome as:

- **pooled** — the outcome was reported in at least 2 studies and a **meta-analysis was conducted** and there is a combined estimate of effect
- **not pooled** — the outcome was reported in at least 2 studies but a **meta-analysis was not conducted** and there is no combined estimate of effect (*e.g.* due to heterogeneity); in this case you may want to give a range of the results of studies
- **not measured** — the outcome was not measured in individual studies
- **not reported** — individual studies may indicate, that the outcome was measured but they did not report the results
- **single study** — there is only one study available (GRADEpro handles this status similar to "pooled", however this category was separated for logical resons (one cannot pool the results of a single study)



Note
It is also important to **provide an explanation** in a footnote or comment describing why studies were not pooled.

> *For authors of systematic reviews*
>
> When the data are imported from a Review Manager 5 file all imported outcomes have a status of being **pooled**.

6.4.5 Rating study limitations

Rating Limitations in Study Design or Execution (Risk of Bias)

You should assess if the studies had limitations in design or execution that were serious enough to downgrade the quality of evidence for this outcome.

To rate study limitations:

- If you think any limitations were **negligible** choose **no**
- If you think there were **serious** limitations choose **serious**
  «this will downgrade the quality of evidence for this outcome by **1 level**»
- If you think there were **very serious** limitations choose **very serious**
  «this will downgrade the quality of evidence for this outcome by **2 levels**»
- Explain your choice in a footnote whenever you downgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Authors of Cochrane Collaboration systematic reviews, can use tools provided in chapter 8 of the Cochrane Handbook to assess risk of bias. It provides a detailed discussion of study-level assessments of risk of bias in the context of a Cochrane review and proposes an approach to assessing the risk of bias for an outcome across studies as »low risk of bias«, »unclear risk of bias« and »high risk of bias«. These assessments may be used directly to inform the assessment of study limitations in the GRADE approach. In particular:

- **low risk of bias** would indicate **"no serious limitations"**
- **unclear risk of bias** would indicate either **"no serious limitations"** or **"serious limitations"**
- **high risk of bias** would indicate either **"serious limitations"** or **"very serious limitations"**.

Authors of Cochrane systematic review must use their judgment to decide between alternative categories, depending on the likely magnitude of the potential biases.

Related topics

- Detailed description of study limitations in the GRADE system
- Chapter 8 of the Cochrane Handbook "Assessing risk of bias in included studies"

6.4.6 Rating inconsistency

Rating Inconsistency in Results

You should assess if the results were consistent across studies and if any inconsistency may have been serious enough to downgrade the quality of evidence for this outcome.

To rate inconsistency:

- If you think any inconsistency was **negligible** choose **no**
- If you think there was **serious** inconsistency choose **serious**
  «this will downgrade the quality of evidence for this outcome by **1 level**»
- If you think there was **very serious** inconsistency choose **very serious**
  «this will downgrade the quality of evidence for this outcome by **2 levels**»
- Explain your choice in a footnote whenever you downgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Related topics

- Detailed description of inconsistency in the GRADE system

6.4.7 Rating imprecision

Rating Imprecision of Results

You should assess if the results are precise enough and if any imprecision of the results may have been serious enough to downgrade the quality of evidence for this outcome. Imprecision is defined differently for authors of systematic reviews and for guideline panels.

To rate imprecision:

- If you think the results were precise choose **no**
- If there was **serious** imprecision choose **serious**
  «this will downgrade the quality of evidence for this outcome by **1 level**»
- If there was **very serious** imprecision choose **very serious**
  «this will downgrade the quality of evidence for this outcome by **2 levels**»
- Explain your choice in a footnote whenever you downgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Related topics

» Detailed description of imprecision in the GRADE system

6.4.8 Rating publication bias

Rating Publication Bias

You should assess if there is a probability of publication bias and if reporting bias may have been serious enough to downgrade the quality of evidence for this outcome.

To rate probability of the publication bias:

- If you think there is no evidence of publication bias choose **unlikely**
- If there is **high** probability of publication bias choose **likely**
  «this will downgrade the quality of evidence for this outcome by **1 level**»
- If there is **very high** probability of publication bias choose **very likely**
  «this will downgrade the quality of evidence for this outcome by **2 levels**»
- Explain your choice in a footnote whenever you downgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Related topics

» Detailed description of publication bias in the GRADE system

6.4.9 Rating magnitude of effect

Rating Magnitude of the Effect

You should assess if the effect was large or very large and, if so, upgrade the quality of evidence accordingly for this outcome.

To rate magnitude of the effect:

- If the effect was not large (RR between 0.5 and 2.0) choose **no**
- If the effect was **large** (RR either >2.0 or <0.5 based on consistent evidence from at least 2 studies, with no plausible confounders) choose **RR >2 or <0.5**
  «this will upgrade the quality of evidence for this outcome by **1 level**»
- If the effect was **very large** (RR either >5.0 or <0.2 based on direct evidence with no major threats to validity) choose **RR >5 or <0.2**
  «this will upgrade the quality of evidence for this outcome by **2 levels**»
- Explain your choice in a footnote whenever you upgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Note
A relative risk (RR) of 0.5 corresponds to a relative risk reduction (RRR) of 50% and a RR of 0.2 corresponds to a RRR of 80%.
When the events in the control group are not frequent, OR and HR can be assumed to be equal to the RR for the application of this criterion.

Related topics

» Detailed description of the magnitude of the effect in the GRADE system

6.4.10 Rating dose-response gradient

Rating the Dose-Response Gradient

In **randomised trials** and in **observational studies downgraded for any reason** do not rate the presence of dose-response gradient and choose **no**

You should assess if there was a dose-response gradient only in **observational studies not downgraded for any reason**. If a dose-response gradient was present, upgrade the quality of evidence for this outcome.

To rate the presence of dose-response gradient:

- If there is no evidence of dose-response gradient choose **no**
- If there is evidence of dose-response gradient choose **yes**
  «this will upgrade the quality of evidence for this outcome by **1 level**»
- Explain your choice in a footnote whenever you upgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Related topics

» Detailed description of the dose-response gradient in the GRADE system

6.4.11 Rating influence of all residual confounding

Rating the Influence of All Plausible Residual Confounding

In **randomised trials** and in **observational studies downgraded for any reason** do not rate the influence of all plausible residual confounding and choose **no**

Only in **observational studies not downgraded for any reason** you should assess if the influence of all plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results show no effect. In either of these two cases, upgrade the quality of evidence for this outcome.

To rate the effect of all plausible residual confounding:

- If there is no evidence that the influence of all plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results show no effect choose **no**
- If there is evidence that the influence of all plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results show no effect choose **yes**
  «this will upgrade the quality of evidence for this outcome by **1 level**»
- Explain your choice in a footnote whenever you upgrade the quality of evidence for any reason, because it is important for others to understand your choice.

Related topics

» Detailed description of the influence of all plausible residual confounding in the GRADE system

6.5 Completing the summary of findings for an outcome

Completing the Summary of Findings for an Outcome

The summary of findings contains the information about the **usual burden of the outcome** (control risk or control score) and the **magnitude and precision of the estimated effect**.

For each outcome, it is necessary to provide the results from the meta-analyses or single studies. These results are entered and presented differently depending on whether the outcome is dichotomous or continuous. The editing pane for the Summary of Findings, therefore, appears differently for each type of outcome.

Depending on whether the outcome is dichotomous or continuous, results are entered and presented differently and the editing pane appears differently for each type of these outcomes.

In general, the data necessary to complete the summary of findings include:

1. description of the outcome, how it was assessed and in what time period
2. number of participants in the studies
3. assumed risk either for dichotomous or continuous outcome
4. estimate of relative effect for dichotomous outcome or of the effect for continuous outcome
5. corresponding risk for a dichotomous outcome

6. Comments

When importing data from a Review Manager 5.0 file different information will be imported from the meta-analysis for dichotomous and continuous outcomes.

Detailed description for data entry and editing for

- dichotomous outcomes
- continuous outcomes

6.5.1 Dichotomous outcomes

Completing Information about Dichotomous Outcomes

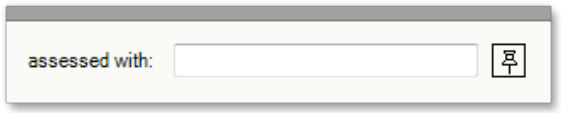
You will need to enter or edit (depending on whether data were imported from Reviw Manager 5.0 file) the following information for a dichotomous outcome:

- how the outcome was assessed
- length of follow-up
- number of participants
- assumed risk (baseline risk)
- relative effect
- corresponding risk

6.5.1.1 Providing the outcome measure

Providing Information on How the Outcome was Assessed

You may provide additional information on how the outcome was assessed by typing it in the box.



The specific name of the measurement tool or a description of how the outcome was measured (*e.g.* 'skeletal event' was defined as: fracture, surgery, or compression) will always need to be entered regardless of whether data were imported from a Review Manager 5.0 file or not.

6.5.1.2 Length of Follow-up

Specifying the Length of Follow-up

You can provide information on the length of the follow-up in the studies that reported each outcome. This is important, since the interpretation of the observed effect depends on the time in which the events occurred.

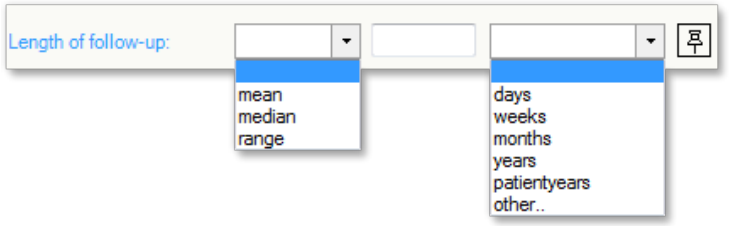
You will have to make **judgments across studies**, because it is unlikely that the duration of follow-up was the same in all studies.

Note
The **duration of intervention** (which should be described in the intervention type, *e.g.* treatment with glucosamine for 8 weeks) may differ from the **duration of follow-up** (*e.g.* assessment of health related quality of life or function after 1 year).

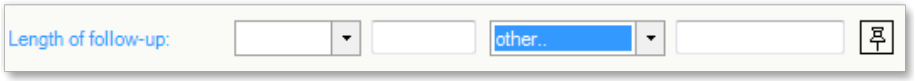
You should provide the information on the length of follow-up for each outcome separately since they may have been measured in different periods.

To provide the information on the time of follow-up:

- Choose **mean**, **median**, or **range** from a drop-down list
- Type a **number** (for mean or median) or a range (e.g. **4 to 7**) in the box
- Choose the unit of time from a drop-down list (**days**, **weeks**, **months**, **years**, **patient-years**, or **other...**)



If you choose other... measure of time an additional box will appear allowing you to specify the unit (*e.g.* hours, etc.)



For authors of systematic reviews

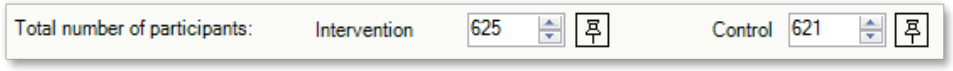
The length of follow-up at which time the outcome was measured will always need to be entered regardless of whether data were imported from a Review Manager 5.0 file or not.

6.5.1.3 Participants

Providing Number of Participants

You will need to enter a total number of participants in control groups and intervention/treatment groups in the studies that contribute to the results. Usually this information can be obtained directly from a meta-analysis graph (forest plot).

For a **continuous outcome** you should specify **total number of participants** in the intervention/treatment and control groups.



For a **dichotomous outcome** you should also specify the **number of participants (patients) in whom the event occurred**.

In the **grey boxes to the right** you will see the **percentage** of total subjects in whom the outcome occurred. The check-box next to the control group event rate lets you choose this rate as a basis for calculation of the absolute effect observed in the studies.

For authors of systematic reviews

When **importing** from Review Manager 5.0 file this information will be **automatically entered** into GRADEpro.

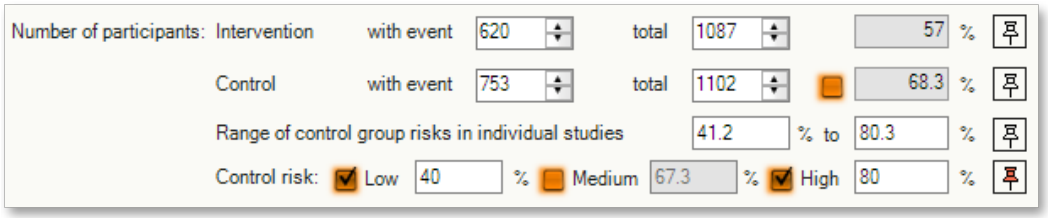
6.5.1.4 Assumed risk

About Assumed Risk

The results presented in the evidence (Summary of Findings) table are built around the assumption of a **consistent relative effect**. It is therefore important to consider the **implications of this effect for populations at different baseline risks**. Once the assumed risk(s) are chosen, a corresponding risk after an intervention (intervention group risk) can be calculated using the meta-analytic risk ratio or odds ratio.

You can provide **up to three typical (assumed, control or baseline) risk values** out of **four possibilities** provided in GRADEpro. There are four checkboxes (here colored orange for emphasis) next to fields where assumed risks may be entered.

Below you will find few suggestions how to choose assumed risk(s) and how to present it.



It is important to indicate with a footnote why a particular assumed risk was chosen.

In the above example two assumed risks were provided: low and high. These will be shown in a Summary of Findings table as two separate populations at different baseline risk: low and high, and the corresponding risk will be calculated for each of the assumed risks based on the relative effect (presumed to be constant irrespective of baseline risk).

bisphosphonates added to anti-cancer treatment for women with advanced breast cancer and bone metastases

Patient or population: women with advanced breast cancer and bone metastases
Settings: ???
Intervention: bisphosphonates added to anti-cancer treatment[1]

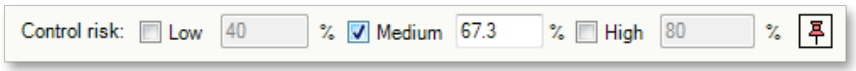
Outcomes	Illustrative comparative risks* (95% CI) Assumed risk Control	Corresponding risk bisphosphonates added to anti-cancer treatment	Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
Skeletal events (new bone metastases, fractures, spinal cord compression, irradiation or surgery, bone pain) (follow-up: 1-3 years)	**Low risk population**[2]		RR 0.83 (0.78 to 0.89)	2189 (8)	⊕⊕⊕○ **moderate**[3,4]	12 studies reported event rates: mean reduction 28% (14 - 48). 10 reported significant reductions.
	40 per 100	33 per 100 (31 to 36)				
	High risk population[2]					
	80 per 100	66 per 100 (62 to 71)				

Choosing the assumed risk to present in the evidence (SoF) table

The assumed risk can be derived from a variety of populations at different baseline (control) risk of an outcome and at different lengths of follow-up. Ideally, risks would reflect groups that clinicians can easily identify on the basis of their presenting features.

Suggestions:

- present the **median control group risk from the studies** included in a meta-analysis.
  If there is little variation in the baseline risks across the studies included in the meta-analysis, you may calculate the median control group risk across studies. This median risk across studies is automatically calculated when data is imported from Review Manager 5.0 file and inserted in the field labelled: **Medium**.



- present up to 3 risks based on the **control group risks in the studies** included in the meta-analysis. You can calculate a low, medium and high assumed risk from the studies. Alternatively, for a high and low risk population you can choose the **second highest** and **second lowest** control group risks in the included studies.

- present a baseline risk from **representative observational studies**. You may enter it as a low, medium, or high risk depending on your judgement, and explain how the assumed risk was obtained.
- present a **mean baseline risk from the studies** included in meta-analysis calculated as number of patients in the control groups with event divided by a total number of patients in control groups. This is automatically calculated by GRADEpro and displayed after the total number of subjects in control groups. This is a natural choice when there is only one study available. Please

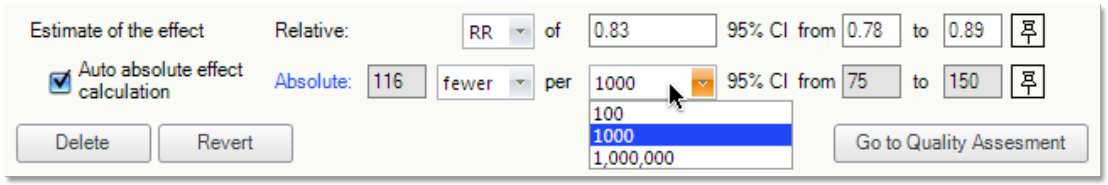
  Note

  This option is prone to bias as it will not account for differences among the studies and it ignores weight of the studies. Ideally a weighted mean should be used to represent this mean. This can be done by calculating a weighted mean and showing this in the medium risk group.

Whichever assumed risk you choose to present you should explain your choice in a footnote.

Choosing the denominator to present risks

You can present the assumed risk and corresponding risk (in a Summary of Findings table) and the absolute effect of an intervention (in GRADE evidence profile) as number of subjects per 100, per 1000, or per 1,000,000. We suggest that, **by default, risk is presented per 1000 people**. You can choose to present risk per 1,000,000 if the events are rare or per 100 if the events are more frequent.



Providing range of control-group risks in individual studies

Information about the range of control-group risks in individual studies is provided because it allows:

- choosing the control group risk(s) that will appear in the Summary of Findings table (it serves as a step in completion of an evidence table, but is not shown in the Summary of Findings table)
- in case that only one (*e.g.* median) control group risk is chosen to be displayed, information about the range of control-group risks in individual studies may be used as a reference, however, it is less important when creating Summary of Findings tables for systematic reviews, because this information is available in the review itself

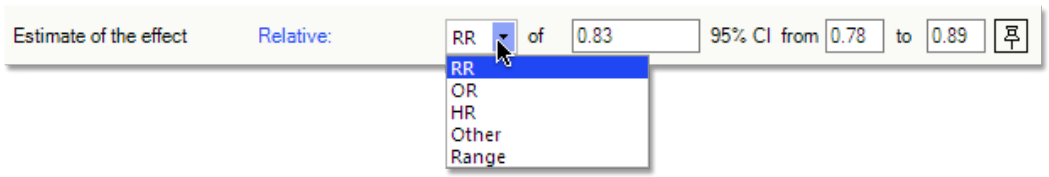
6.5.1.5 Relative effect

About the Relative Effect

The relative effect for a dichotomous outcome from a single study or a meta-analysis will typically be a risk ratio (relative risk), odds ratio, or occasionally a hazard ratio.

Authors of Cochrane Reviews who develop a SoF table, should present the relative effect measure they used in the meta-analysis. However, other users may want to present a different relative effect measure than what they observe in the literature they use to develop a SoF table or GRADE profile (see below for how to convert the relative effect from the meta-analysis to a different relative effect). You can select the relative effect used in the meta-analysis or those of one or more studies if no pooled estimate is available from the drop down menu. There are 5 options:

- RR (Risk Ratio or Relative Risk)
- OR (Odds Ratio)
- HR (Hazard Ratio)
- Other
- Range (when meta-analysis was not conducted and a range of relative effects from studies can be presented; you should specify the type of effect measure that was used, *e.g.* RR or OR)



You can enter the point estimate and the confidence limits (a range of point estimates if there were several studies but no meta-analysis).

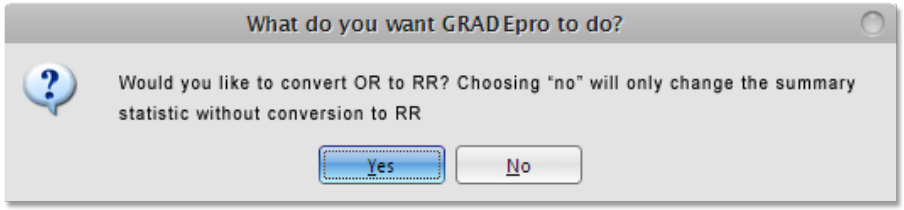
You should add footnotes to provide more information about the estimate of the effect.

When results are not pooled and a range of effects cannot be presented, you can briefly summarize the qualitative assessment of the pattern of the results in the comments section for this outcome. However, you should avoid "vote counts" — reporting numbers of "positive" and "negative" studies, because they are not informative.

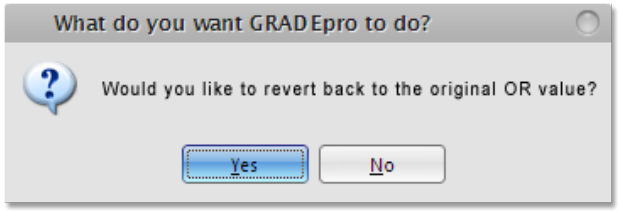
If you imported data from a Review Manager 5.0 file, the point estimate and confidence limits for a relative effect will be automatically imported.

Note

GRADEpro can **convert an Odds Ratio or Hazard Ratio** (which was previously entered or imported into GRADEpro) **into an RR**. After entering the values for OR or HR select RR from the drop down menu. A pop-up window will include a question whether you would like to convert OR or HR to RR. If you want to convert choose **YES**, if you want just to change the label for summary statistics choose **NO**.



The original OR or HR is saved in GRADEpro so that after converting them to RR, you can revert back to an original OR or HR. Select OR or HR in the drop down menu and choose YES to the pop-up window "Would you like to revert back the original value?"



The formulae GRADEpro uses to convert OR or HR to RR.

6.5.1.6 Corresponding risk

About Corresponding Risk

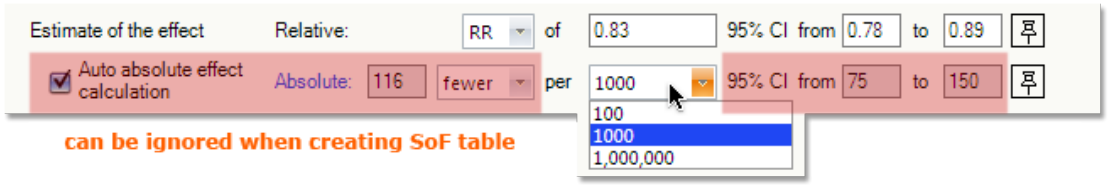
The **corresponding risk** is a measure of the burden of the outcomes after the intervention is applied, *i.e.* the risk of an outcome in people who were exposed or received an intervention. It is based on the relative magnitude of an effect and assumed (baseline) risk.

GRADEpro will **automatically** calculate the corresponding risk based on the assumed risk and the relative effect size. GRADEpro will calculate the corresponding risk once at least one of the assumed risk values is provided and the magnitude of the relative effect (RR, OR or HR) has been entered. Corresponding risk values are automatically entered into the Summary of Findings table.

If you imported data from a Review Manager 5.0 file, you will notice that the relative effect and the corresponding risk will be in the Summary of Findings table as soon as the data is imported from RevMan.

Note
Auto absolute effect calculation option and **calculation of absolute effect** **do not pertain to SoF tables** and can be ignored. This function is used for creating GRADE evidence profiles.

The only option in this row of **auto absolute effect calculation** that you will need to consider creating a SoF table is the drop down menu for choosing **denominator to present risks**. You can present assumed risk and corresponding risk as number of subjects per 100, per 1000, or per 1,000,000. By default, risk is presented **per 1000** people. You can choose to present risk per 1,000,000 if the events are rare or per 100 if the events are more frequent.



6.5.1.6.1 Calculation of corresponding risk

About Calculation of Corresponding Risk in GRADEpro

GRADEpro will automatically calculate the corresponding risk (in SoF table) or absolute effect (in GRADE evidence profile) based on the assumed risk and the relative effect size according to the following formulae (calculations are based on The Cochrane Handbook Chapter 12.5.4.).

Corresponding risk from Risk Ratio or Relative Risk (RR)

Corresponding intervention risk (per 1000 people) = 1000 × ACR × RR
where
ACR — Assumed Control Risk,
RR — Risk Ratio or Relative Risk

Example

The risk ratio or relative risk (RR) from a meta-analysis is 0.83 (95% CI, 0.78 to 0.89),
The assumed risk (ACR) is 40% (400 per 1000 or 0.4).
therefore
Corresponding risk = 1000 × 0.4 × 0.83 = 334 per 1000.

Corresponding Risk from Odds Ratio (OR)

The odds ratio (OR) is first converted to risk ratio (RR) and the corresponding risk is then calculated as above.
The formula for the conversion of odds ratio to risk ratio uses control event rate (CER) from the meta-analysis:

$$RR = \frac{OR}{1 - CER \times (1 - OR)}$$

where
CER — Control Event Rate (CER = number of people with event in control groups ÷ total number of people in control groups)
OR — odds ratio
RR — Risk Ratio

Example

The odds ratio (OR) from a meta-analysis is 0.64 (95% CI, 0.47 to 0.89),
CER = 112 ÷ 438 = 25.6% or 0.256
therefore
RR = 0.64 ÷ (1—0.256 × (1—0.64)) = 0.70
The assumed risk (ACR) is 10% (100 per 1000 or 0.1).
therefore
Corresponding risk = 1000 × 0.1 × 0.70 = 70 per 1000.

Corresponding Risk from Hazard Ratio (HR)

The hazard ratio (HR) is first converted to risk ratio (RR) and the corresponding risk is then calculated as above from RR.
The formula for the conversion of hazard ratio to risk ratio uses control event rate (CER) from the meta-analysis:

$$RR = \frac{1-e^{HR \times \ln(1-CER)}}{CER}$$

where
CER — Control Event Rate (CER = number of people with event in control groups ÷ total number of people in control groups)
HR — hazard ratio
RR — Risk Ratio

6.5.1.7 Comments

About Comments Column in Summary of Findings Table

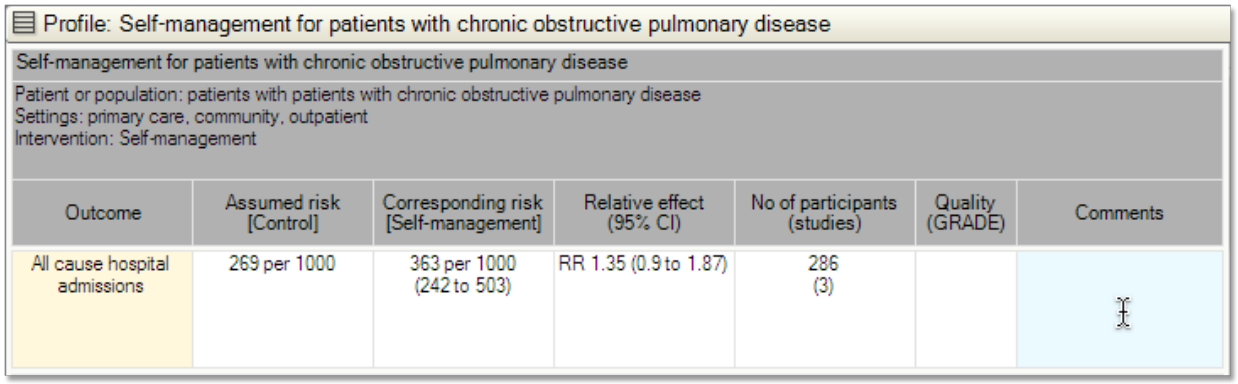
In **Comments** field you can provide additional information to facilitate interpretation of the information about an outcome.

For example, you can provide information about the validity of the outcome measure, patient-importance of an observed effect, or the presence of variables that were associated with the magnitude of effect. You should also flag any important caveats about the results if they occured. Not all rows (outcomes) will need comments, so it is best to leave this field blank if there is nothing warranting a comment.

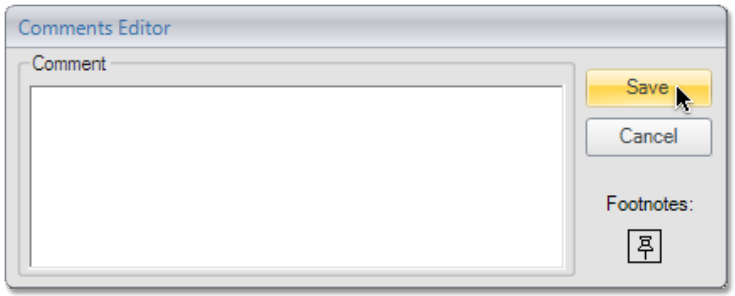
You can also provide additional information about continuous outcomes.

To add a comment

- click on the field for **Comments** in the Summary of Findings table preview pane



a Comments editor will pop-up



- add comments
- add a footnote , if necessary
- click **SAVE**

Note

You will **not** be prompted to add comments.

6.5.2 Continuous outcomes

Completing Information about Continuous Outcomes

The term "continuous" in statistics conventionally refers to data that can take any value in a specified range (Cochrane Handbook Section 9.2.3). Examples of truly continuous data are distance, area and volume. In practice, one can use the same statistical methods for other types of data, most commonly measurement scales and counts of large numbers of events (Cochrane Handbook Section 9.2.4).

The mean difference (MD) or standardised mean difference (SMD) are the two key pooled estimates of effect for continuous outcomes measured on a continuous or ordinal scale. The presentation of the results for these outcomes depends on the type of pooled estimate calculated.

Note

When completing the information for continuous outcomes, it is important to consider the differences in scores that you will present in the Summary of Findings table. You can present the difference in scores between the control group/assumed risk and the intervention/corresponding risk group for

- change in scores from baseline to end of study (change scores)
- scores from the end of the study (final values).

When naming the outcome indicate if it is considered an absolute value or a change (score).
For example, specify **Pain** versus **Change in Pain** or **Quality of Life** versus **Change in Quality of Life**.

Other informative ways of presenting continuous outcomes are the **ratio of means** (*e.g.* the ratio of the mean in weight gain in one group compared to another). A special case that can be considered a continuous outcome is a **comparison of rates or rate ratios** (*e.g.* the number of disease exacerbations per patient or the number of new polyps per patient in one group compared to another).

You will need to enter or edit (depending on whether data were imported from Reviw Manager 5.0 file) the following information for a continuous outcome:

- length of follow-up
- number of participants
- assumed risk (final values or change score in control group)
- estimate of the effect
- measurement scale
- comments

6.5.2.1 Length of Follow-up

Specifying the Length of Follow-up

You can provide information on the length of the follow-up in the studies that reported each outcome. This is important, since the interpretation of the observed effect depends on the time in which the events occurred.

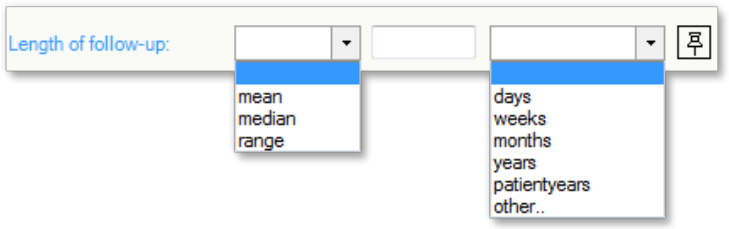
You will have to make **judgments across studies**, because it is unlikely that the duration of follow-up was the same in all studies.

Note
The **duration of intervention** (which should be described in the intervention type, *e.g.* treatment with glucosamine for 8 weeks) may differ from the **duration of follow-up** (*e.g.* assessment of health related quality of life or function after 1 year).

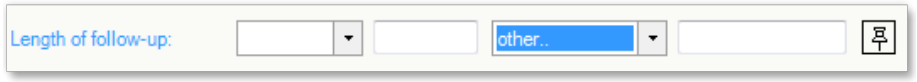
You should provide the information on the length of follow-up for each outcome separately since they may have been measured in different periods.

To provide the information on the time of follow-up:

- Choose **mean**, **median**, or **range** from a drop-down list
- Type a **number** (for mean or median) or a range (e.g. **4 to 7**) in the box
- Choose the unit of time from a drop-down list (**days**, **weeks**, **months**, **years**, **patient-years**, or **other...**)



If you choose other... measure of time an additional box will appear allowing you to specify the unit (*e.g.* hours, etc.)



For authors of systematic reviews

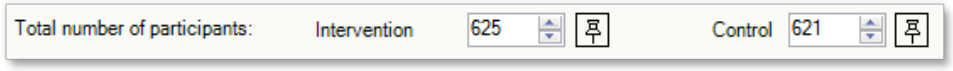
The length of follow-up at which time the outcome was measured will always need to be entered regardless of whether data were imported from a Review Manager 5.0 file or not.

6.5.2.2 Participants

Providing Number of Participants

You will need to enter a total number of participants in control groups and intervention/treatment groups in the studies that contribute to the results. Usually this information can be obtained directly from a meta-analysis graph (forest plot).

For a **continuous outcome** you should specify **total number of participants** in the intervention/treatment and control groups.



For a **dichotomous outcome** you should also specify the **number of participants (patients) in whom the event occurred**.

In the **grey boxes to the right** you will see the **percentage** of total subjects in whom the outcome occurred. The check-box next to the control group event rate lets you choose this rate as a basis for calculation of the absolute effect observed in the studies.

For authors of systematic reviews

When **importing** from Review Manager 5.0 file this information will be **automatically entered** into GRADEpro.

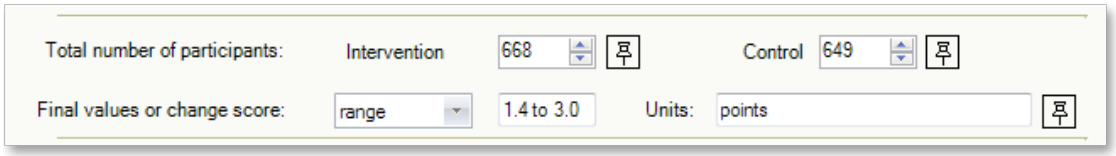
6.5.2.3 Assumed score

About Assumed Risk (Score)

The Assumed Risk (Score) is the score of the participants who received the control intervention (in the context of systematic review or single study) or for whom the recommendation is intended (in the context of clinical practice guidelines).

Unlike dichotomous outcomes that are measured once at the end of a study, continuous variables are measured both at the beginning (baseline – before observation started or interventions were administered) and at the end of a study. Therefore, continuous outcomes can be expressed either as a **change in scores from baseline to end of a study** (*i.e.* change score) or as **final measurements** at the end of the study (*i.e.* final values). **The Assumed Risk (Score) is either a final value or change score in the control group.** Choice of the presentation of control score in the evidence table will depend on the type of data used in the meta-analysis. You should specify in the outcome name or in a footnote which of the two presentations was used (*e.g.* pain intensity *versus* change in pain intensity).

You can estimate the Assumed Risk (Score) by assessing typical scores in different patient groups or at different lengths of follow-up. Ideally, these groups would reflect patients that clinicians can easily identify on the basis of their presenting features. You can derive the scores in the controls either directly from a systematic review from which you obtained the estimate of the intervention effect or you can extract it from observational studies of patients similar to those for whom the intervention is intended.



---

1. Assumed risk (score) when an effect was expressed as a MEAN DIFFERENCE (MD)

You can provide a range, median or mean of the final values or change scores in the control groups.

- range — it can be the highest and the lowest estimate of the scores in the control groups
  Note: **if there are "outliers"** another approach for obtaining high and low estimates would be to use the **second lowest and second highest** control score from the studies.
- median — the mid level score can be based on the median of the scores in the control groups across studies in the systematic review or on data from representative observational studies
- mean — the mid level risk can be based on the weighted mean of the scores in the control groups across studies in the systematic review or on data from representative observational studies.

You should provide a footnote clarifying the source of the assumed risk (score) used.

You should also provide the **units of the scale on which the scores were measured** (*e.g.* kg, points, days, etc.). These units will appear in the evidence profile (SoF table).

Note
Final values or change scores will **not** be automatically imported into GRADEpro from Review Manager 5 file.

---

2. Assumed risk (score) when an effect was expressed as a STANDARDIZED MEAN DIFFERENCE (SMD)

There are three alternative methods for presenting a SMD (» Re-expressing SMD) and therefore three different presentations of final values or change scores in the control groups.

1. when using rules of thumb for effect sizes — **you do not fill in information about the final values or change scores**, you present only the SMD with the corresponding confidence interval
2. when transforming to odds ratio — if you "dichotomize" results of continuous outcome by transforming SMD to OR (or RR), you can apply these numbers to an assumed risk similar to the calculations for dichotomous outcomes. You will need to change the type of the outcome from continuous to dichotomous on the Quality Assessment editing pane and then enter the data according to presentations for dichotomous outcomes. If you choose this option, you can also provide up to 3 risk estimates for different risk groups
3. when using a familiar instrument — for the "back translation" of a SMD to a familiar instrument you can present a mean, median, or range of final values or change scores similar to the presentation for MD (see above). The mean, range or median, however, will be from the studies that used the same familiar instrument.

Note
All of these alternative methods for presenting a SMD have limitations since they require several statistical assumptions that do not hold under all circumstances. However, these approximations facilitate interpretable presentations of results.

---

3. Assumed risk (score) when an effect was expressed as a RATIO OF MEANS

You do not fill in information about the final values or change scores in control groups.

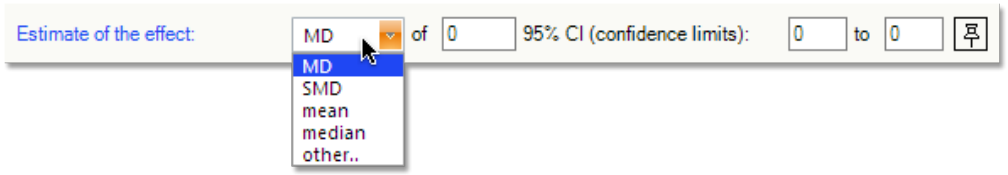
6.5.2.4 Estimate of the effect

About Estimate of the Effect

The estimate of the effect can be presented in a variety of ways. There are 5 options:

- mean difference (MD)
- standardised mean difference (SMD)
- mean
- median
- other

You can select the effect used in the meta-analysis from the drop down menu.



---

1. When using a MD, mean, or median

You can enter the point estimate and the confidence intervals. You can add footnote to provide more information about the effect and confidence intervals as necessary.

> *FOR AUTHORS OF SYSTEMATIC REVIEWS*
>
> If you imported data from a Review Manager 5 file the estimate and confidence intervals will be automatically imported.

---

2. When re-expressing SMDs

You will need to enter or edit the data depending on how the SMD is re-expressed.

1. when using rules of thumb for effect sizes — you should select MD from the drop down menu and fill in the **MD** and confidence intervals. This information is presented in the corresponding risk column of the outcome. See below for suggestions for comments. A footnote should be provided.
2. when transforming to odds ratio — you should calculate OR from the SMD and then apply the OR as with dichotomous outcomes. You will need to change the type of outcome to **dichotomous** on the Quality Assessment editing pane and then enter the data according to presentations for dichotomous outcomes. You should include the original SMD value in the comments section (*e.g.* "Numbers estimated using a standardised mean difference of XX (95% CI YY to ZZ)")
3. when using a familiar instrument — for the "back translation" of an SMD to a familiar instrument you should choose a representative study and then calculate the difference by multiplying the SMD by the SD (standard deviation) of the mean change in the control group in this study. You should select **MD** as the type of estimate and then enter the difference and the

confidence intervals. These numbers will be displayed in the assumed risk and corresponding risk columns similar to the MD presentation.

Note

If you import data from a Review Manager 5 file you will need to select **MD** in the drop down menu and then change the numbers that were originally imported as SMD into the difference and confidence intervals calculated using data from representative study.

---

3. When using the ratio of means

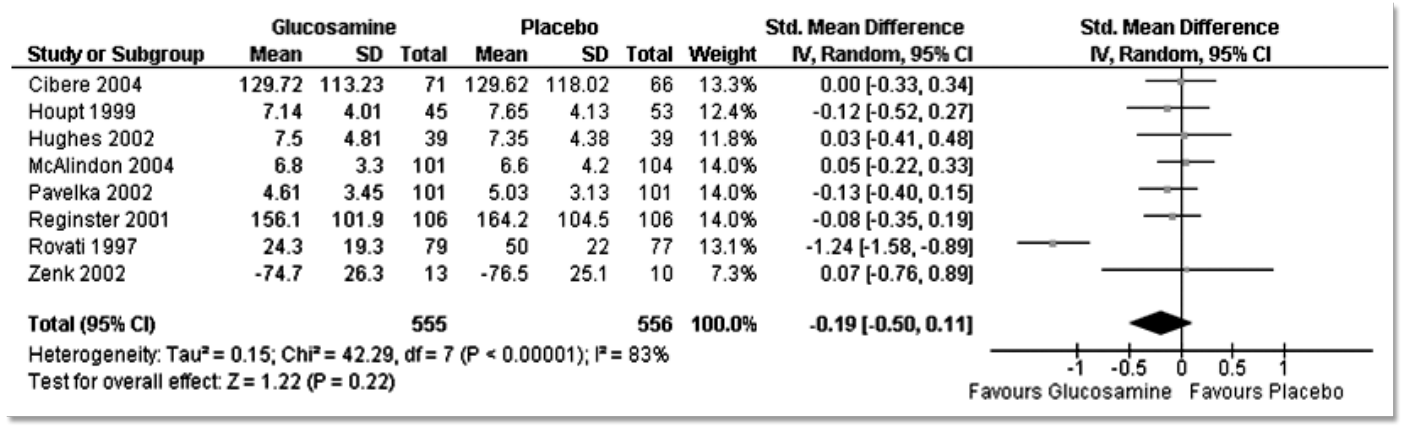
The ratio of means represents the weighted average of the mean scores in the intervention group divided by mean score in comparison group. Only the relative effect is given and the final values or change score in control groups or in the intervention group are not given. You should choose **other...** from a drop-down menu and enter information about the ratio of means directly into the comments column.

6.5.2.4.1 Re-expressing SMD

About Re-expressing SMD

When studies have used **different instruments** to measure the same construct, a standardized difference in means (SMD) may be used in meta-analysis for combining continuous data. The SMD expresses the intervention effect in **standard deviation (SD) units** rather than the original units of measurement. The mean difference (whether change from baseline to end of study, or end of study value) is standardized by dividing it by the standard deviation in the control group in this study. The standardized means from the individual studies are then combined in meta-analysis to calculate SMD. Consequently, the value of SMD depends on both the **size of the effect** (the difference between means) and the standard deviation of the outcomes (the inherent **variability among participants**).

Forest plot used in examples below.



There are three options for re-expressing the SMD facilitating its interpretability:

---

1. Re-expressing SMDs using rules of thumb for effect sizes

Rules of thumb exist for interpreting SMDs or "effect sizes". If you choose this mode of presenting SMD **you should include a rule of thumb** in the Comments column of a Summary of Findings table. You should bear in mind that some methodologists believe that such interpretations are problematic, because patient importance of a finding is context-dependent and not amenable to generic statements.

Rule of thumb according to Cohen's interpretation of effect size

- 0.2 represents a small effect
- 0.5 represents a moderate effect
- 0.8 represents a large effect

There are variations of Cohen's interpretation. An example might be:

- <0.41 represents a small effect
- 0.40 to 0.70 represents a moderate effect
- >0.70 represents a large effect.

A sample Summary of Findings table presenting SMD from the above example using Cohen's interpretation of effect size

Outcomes	**Illustrative comparative risks*** (95% CI)		**Relative effect (95% CI)**	**Number of participants (studies)**	**Quality of the evidence (GRADE)**	**Comments**
	Assumed risk **With no treatment**	Corresponding risk **With glucosamine**				
Pain Measured with different scales in the different studies. Lower scores mean less pain. (follow–up: 3 months)		The mean pain score in the intervention groups was **0.19 lower** (0.50 lower to 0.11 higher)		1111 (8)	⊕⊕◯◯ **Low**3,4	A standard deviation of 0.2 represents a small difference between groups

---

2. Re-expressing SMDs by transformation to odds ratio

A transformation of a SMD to an odds ratio (OR) is possible. Due to the underlying assumptions to make this conversion, the results are only an approximation. To calculate OR use the formula:

$$\ln(OR) = \frac{\Pi}{v3} SMD$$

where Π/v3 is approximately 1.8138

The estimated odds ratio can then be entered similarly as for a dichotomous outcome. The assumed risk (control group risk) refers to the proportion of people who have improved by some unspecified amount (or those without an event) in the continuous outcome ("responders"). GRADEprofiler can then automatically calculate the corresponding risk based on the assumed risk entered and present the results as dichotomous outcome. You should add a comment such as, "numbers estimated using a standardised mean difference of XX (95% CI YY to ZZ)"). If you select this option you will be able to choose more than one assumed risk value as for other dichotomous outcomes.

In the above example the SMD was 0.19 which multiplied by 1.8138 gives 0.34. If ln (OR) = 0.34 then OR = 1.41. The assumed risk was 0.9.

Note

In the SoF below the outcome is number of people who had little or no pain (NOT number of people with pain).

A sample Summary of Findings table presenting SMD from the above example using transformation to odds ratio

Outcomes	**Illustrative comparative risks***	**Relative**	**Number of participants**	**Quality of the**	**Comments**

	(95% CI) Assumed risk **With no treatment**	Corresponding risk **With glucosamine**	**effect (95% CI)**	**(studies)**	**evidence (GRADE)**	
Little or no pain Measured with different scales in different studies (Follow-up: 3 months)	**9 per 100**	**12 per 100** (11 to 20)	**OR 1.41** (1.22 to 2.48)	1111 (8)	⊕⊕◯◯ **Low**	Numbers estimated using a standardised mean difference of -0.19 (-0.50 to 0.11)

---

3. Re-expressing SMDs using a familiar instrument

Because the SMD is based on standardized means from the included studies and not a specific scale, it is unit-less. This makes the interpretation of the effect very difficult. To better understand the effect it can be re-expressed by applying the calculated SMD back into one of the original studies and depicted on the scale used in that study. To back transform the SMD to familiar scale

- select a study included in the original meta-analysis that is representative of the population and intervention and at a low risk of bias
- multiply the standard deviation of the control group (end of study mean or mean change from baseline to end of study) by the pooled SMD

This resulting number is the mean difference (MD) and can be presented in the Summary of Findings table in the format of MD (mean difference) for the scale used in the representative study.

Note

One should interpret such results with caution since back-translation of the effect size is based on the results of only 1 study.

A sample Summary of Findings table presenting SMD from the above example using back-translation to a familiar instrument

Outcomes	**Illustrative comparative risks*** (95% CI) Assumed risk **With no treatment**	Corresponding risk **With glucosamine**	**Relative effect (95% CI)**	**Number of participants (studies)**	**Quality of the evidence (GRADE)**	**Comments**
Pain WOMAC. Scale from 0, no pain, to 20, worst pain (follow –up: 3 months)	The mean pain scores ranged from 5.0 to 7.6	The mean pain score in the intervention group was **0.8 lower** (2.1 lower to 0.5 higher).		1111 (8)	⊕⊕◯◯ **Low**	Scores estimated using a standardised mean difference of -0.19 (-0.50 to 0.11)

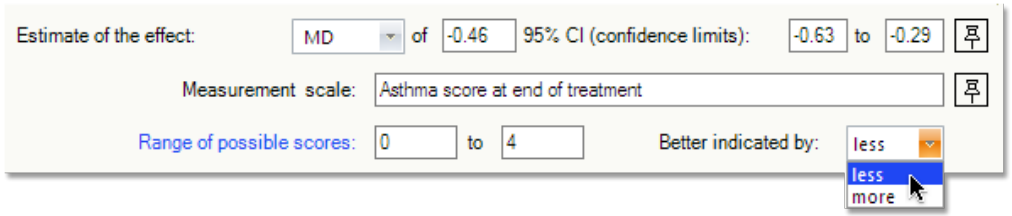
6.5.2.5 Measurement scale

Describing Measurement Scale

Measurement scale

You can enter

- specific name of the measurement scale or the familiar instrument used to re-express SMD (*e.g.* SF-36, WOMAC, etc.)
- description of how the outcome was measured (*e.g.* visual analogue scale)
- statement that the "outcome was measured on different scales in different studies" in case of SMD re-expressed as an effect size



Range of possible scores

You should enter the range of possible scores on the scale used to measure the outcome. You can provide a range of numbers or numbers and text to describe the scores. For example, 0 to 100 or "0 (no difficulty) to 5 (severe difficulty)"

Better indicated by

You can fill in the direction of the effect on the scale. However, this information does not appear in the Summary of Findings table.

> *For authors of systematic reviews*
>
> When importing from a Review Manager 5 file any of the above information is **not imported**.

6.5.2.6 Comments

Specific Comments for Continuous Outcomes

For general discussion on providing comments in SoF table read about comments column in Summary of Findings table.

It is specifically recommended that you use comments for outcomes in which SMD was calculated. Following are suggestions for the 3 options for presenting SMDs

1. when using rules of thumb for effect sizes — the size of the effect (*e.g.* according to Cohen's effect sizes) should be entered into the comments section (*e.g.* "A standard deviation of XX represents a high | moderate | small difference between the groups")
2. when transforming to odds ratio (or other dichotomous outcome measure) — you should include the SMD value in the comments section (*e.g.* "Numbers estimated using a standardised mean difference were XX (95% CI YY to ZZ)")
3. when using a familiar instrument —you should include the SMD value in the comments section (*e.g.* "Numbers estimated using a standardised mean difference were XX (95% CI YY to ZZ)")

7. Managing evidence profiles

Managing Evidence Profiles

You can import save, export, and print any evidence profile.

- Importing data from Review Manager 5.0 file or GRADEpro files
- Copying, duplicating, and moving
- Exporting Summary of Findings table and importing it into Review Manager 5
- Exporting to Microsoft® Office Word
- Exporting as picture
- Exporting as HTML
- Printing

7.1 Importing

Importing data to create evidence tables

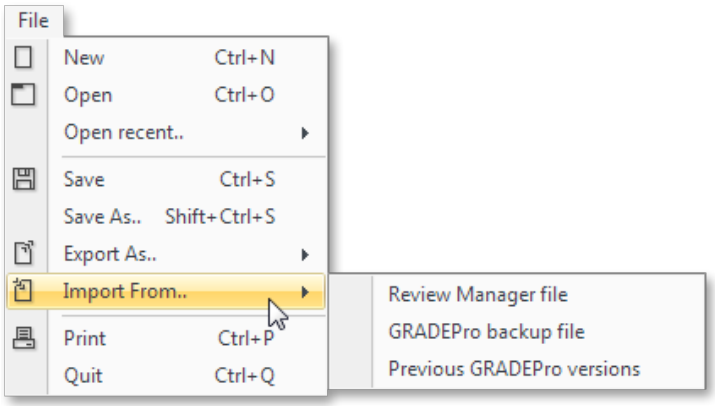
You may import data to GRADEpro from the following:

- Review Manager version 5.0 or later data file (*.rm5)
- GRADEpro backup file (*.grd)
- Files created with previous versions of GRADEpro (*.xml)

If you are an author of a **Cochrane systematic review** you will most likely want to import your data from RevMan 5.
If you used previous versions of GRADEpro (1.0 or 2.0) to create GRADE evidence profiles that you would like to review or amend, GRADEpro will convert them for you and let you save them as GRADEpro 3.0 files.

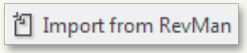
To import data

- Go to **File** menu, choose **Import from...**
- Choose a file type you want to import data from
- Choose the file in a dialog box that opens.



For authors of systematic reviews

There are two ways to start importing data from Review Manager (RevMan) 5.0+:

- Go to **File** menu, choose **Import from...** and then **Review Manager file**
- Click 

on a toolbar

This will open a dialog box in which you should choose the Revman file (*.rm5) you want to import data from.

7.2 Exporting evidence tables

Exporting Evidence Tables

GRADEpro exports groups of profiles, profiles, or outcomes into the followin file formats:

- Summary of Findings (SoF) table (may be imported into RevMan 5.0)
- Microsoft® Office Word document (may be further edited and used in a publication)
- HTML document (may be published online or imported to a word processor)
- Image (available bitmap image formats include: JPEG, GIF, BMP, and PNG)
- XML file

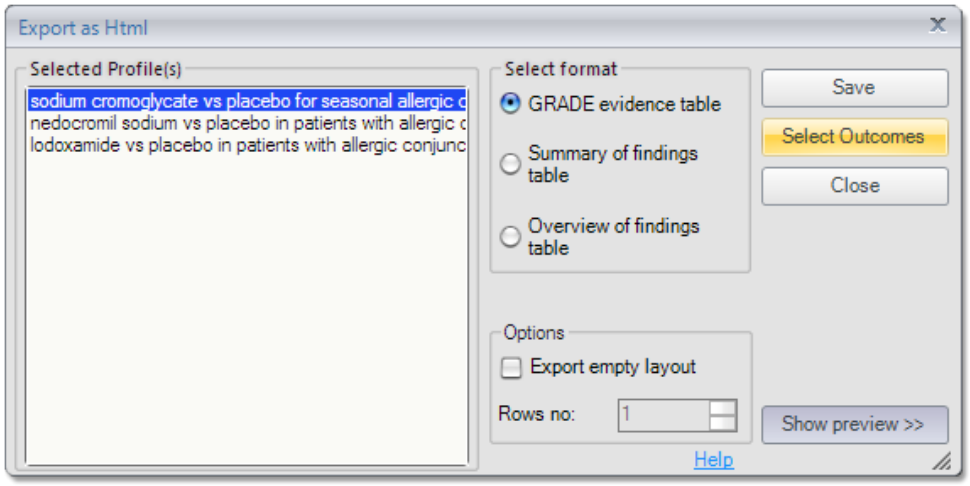
7.2.1 Previewing GRADE evidence profile

Previewing GRADE evidence profile

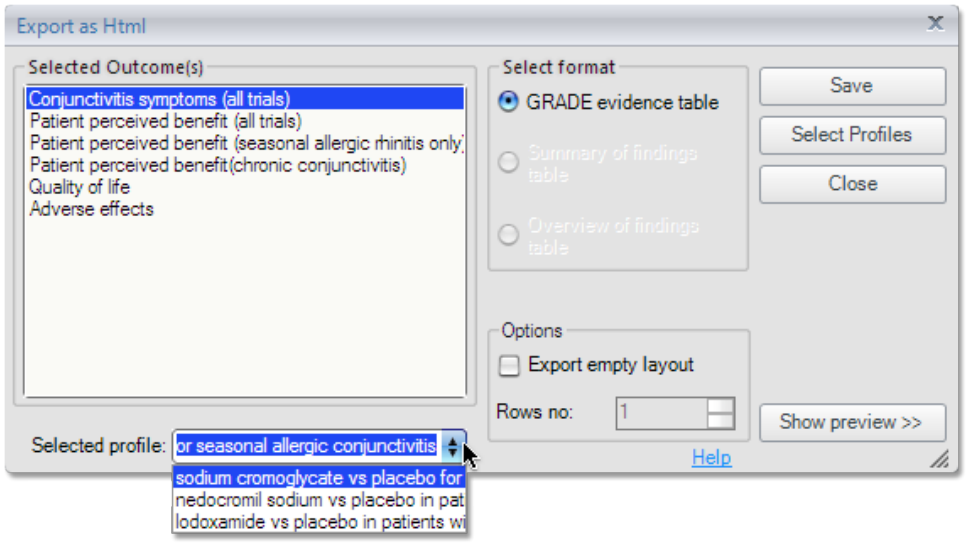
Previewing a GRADE evidence profile (table) is similar to previewing Summary of Findings table.

However, when previewing GRADE evidence profile, **you can select which outcomes to include in the profile**.

- click on the profile you would like to view and then click **Select Outcomes**



- a listo of outcomes will be shown
- you can choose which outcomes to preview or export



- use the drop-down menu to change the profile

7.2.2 Previewing SoF table

About Previewing Summary of Findings Table

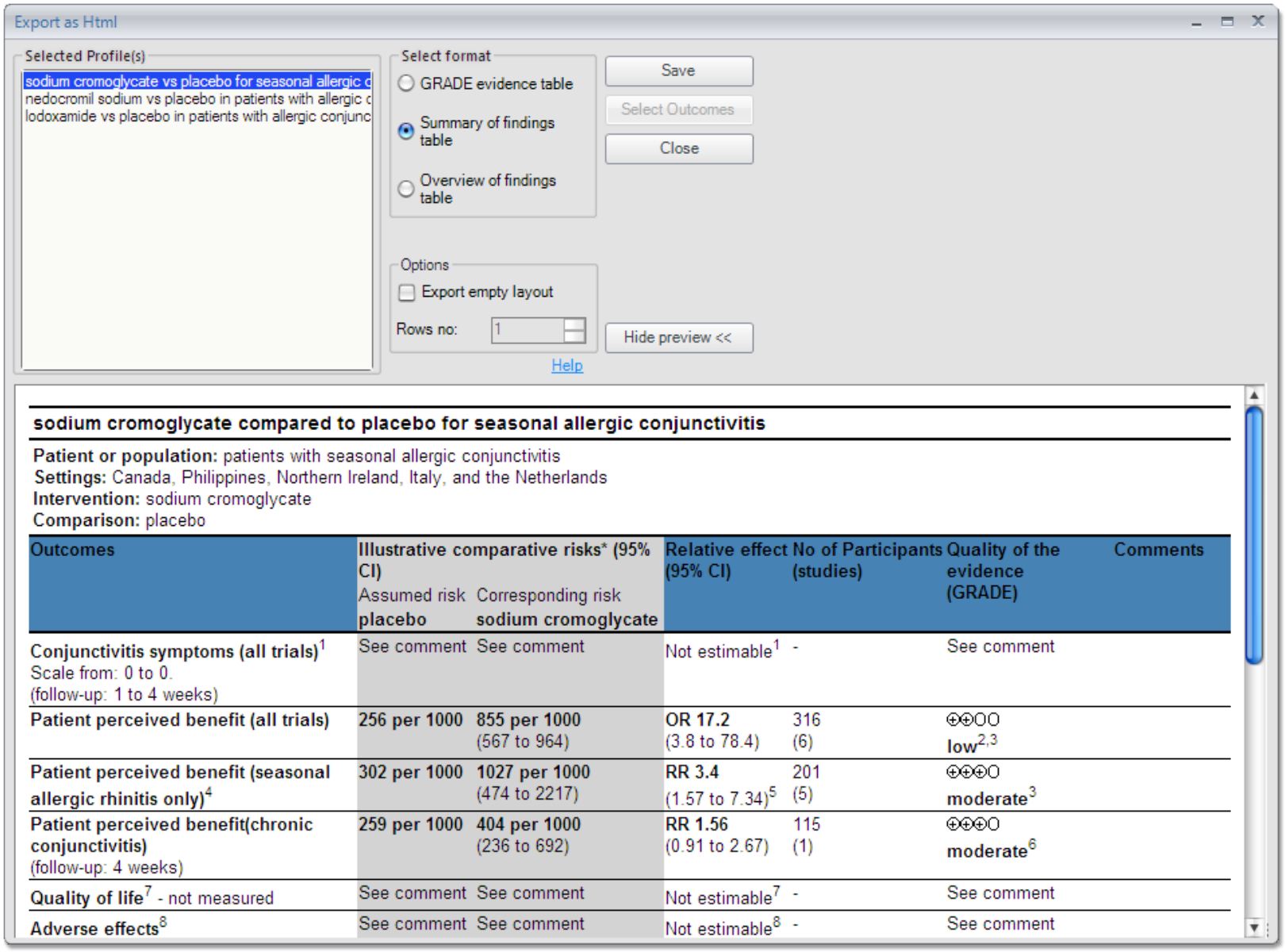
You can **preview a Summary of Findings table** to see **exactly** how your table will appear after it has been saved or exported.

The rows of a **Summary of Findings table** should include all desirable and undesirable outcomes (up to a maximum of 7 outcomes) that are essential for decision-making listed in order of importance. Details of scales and time-frames should be provided. You should aim to decide which outcomes are important for the Summary of Findings table during protocol development and before you undertake the review. However, you should be alert to the possibility that the importance of an outcome (*e.g.* a serious adverse effect) may only become known after the protocol was written or the analysis was carried out, and should take appropriate actions to include these in the Summary of Findings table. You should list these outcomes in the table whether data are available or not.

If there is an excessive number of outcomes in the review, authors will need to omit the less important outcomes. Serious adverse events should be included, but it might be possible to combine minor adverse events, and describe this in a footnote (note that it is not appropriate to add events together unless they are known to be independent). Multiple time points will be a particular problem. In general, to keep the table simple, only outcomes critical to decision making should be presented at multiple time points. The remainder should be presented at a common time point.

Viewing the table in this format after you have edited or entered data in GRADEpro shows you how the table will appear and facilitates detecting errors or formatting issues.

It is especially recommended that you **preview your SoF table before** exporting it to Review Manager.



You can preview a SoF table clicking **Preview SoF table** on the Toolbar. This will open Export as HTML dialog box with a preview enabled.

Summary of Findings table is chosen by **default in the preview** window. However, you have the option of viewing 3 different tables:

1. GRADE evidence table (full GRADE evidence profile)
2. Summary of Findings table (specifically created for Cochrane systematic reviews)
3. Overview of Reviews table (showing results by outcome for multiple comparisons/profiles)

Choosing a SoF to preview

All profiles available in the current GRADEpro file will be listed in the left-hand box. You can choose a profile for which you would like to preview a table. Click on the profile or comparison to view.

You can preview multiple SoF tables for multiple profiles by Ctrl-clicking them or left-clicking and dragging the mouse over the profile names.

Note

An additional feature is available when previewing the GRADE evidence profile — you may select which outcomes to view in the table.

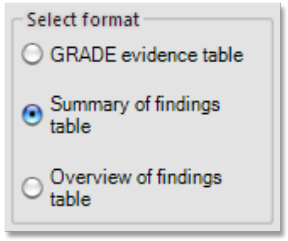
Saving SoF table

You can save your table in HTML format directly from the Preview SoF screen. To export a SoF table to Review Manager *.sof file or save it in different format (Microsoft® Office Word, picture, or XML) you will have to close preview and go to **File » Export as...** .

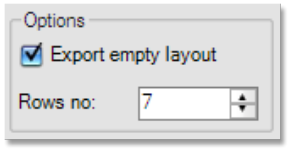
Export Empty Template

On occasion you may wish to save an empty table for later printing or importing it and completing in a word processor. That is, **you can produce an empty template** of a GRADE evidence profile, SoF table or Overview of Reviews table from the preview screen.

- **Choose format** of a table you would like to save



- check **Export Empty Layout**



- specify the number of rows by typing in a number or using the arrow buttons
- **save** the file.

7.2.3 Exporting summary of findings table

Exporting a Summary of Findings Table

Before you export Summary of Findings table to Review Manager

- choose the outcomes you want to include in SoF table in your review (up to 7)
- organise these outcomes by their patient-importance (from the most important to the least important) by rearranging them in the Tree pane

Export SoF table

To include a Summary of Findings table you created in GRADEpro in a Cochrane systematic review you first need to export it in a format suitable for importing into Review Manager 5.0.

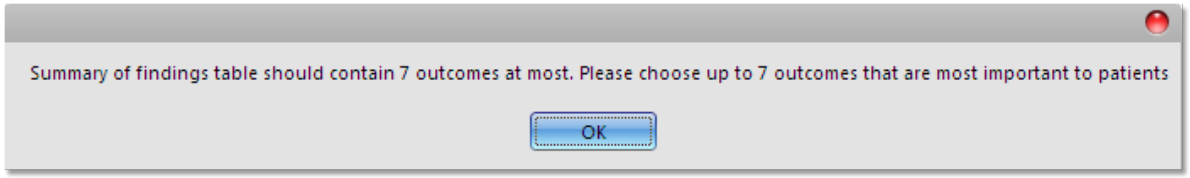
To export SoF table for importing into Revman

- **select profile** (comparison) you would like to export in the Tree of profiles
- from the menu **choose File » Export as... » Review Manager SoF**
- name and save the file

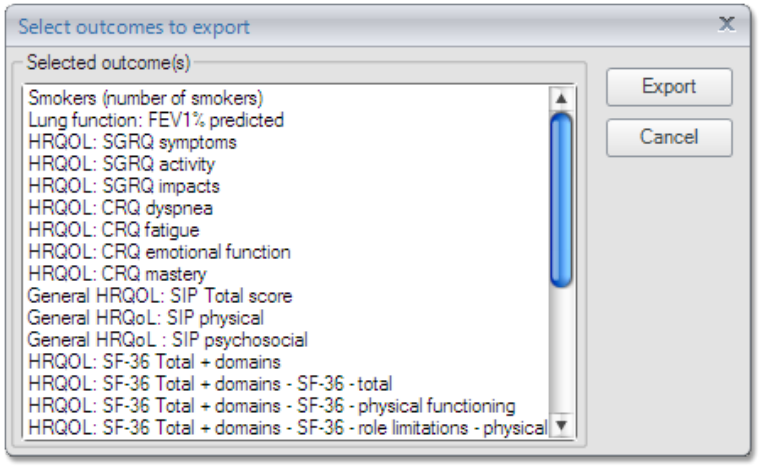
Note
The file will be saved with an extension ***.sof**, that is suitable for importing into RevMan 5.0.

Number of outcomes to export

It is highly recommended that Summary of Findings tables for Cochrane systematic reviews include **no more than 7 outcomes**. If the profile you have chosen includes more than 7 outcomes, a pop-up box will appear warning you that there are too many outcomes.



Upon clicking OK, a list of outcomes will appear.



You will be prompted to choose which outcomes to include. You may select outcomes by holding Control key and clicking them.

7.2.3.1 Importing SoF table into RevMan

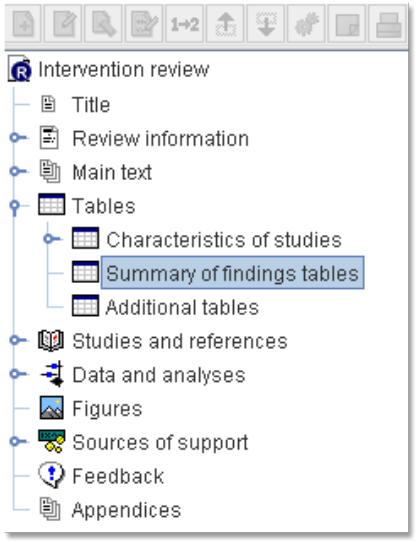
Importing SoF Table into Review Manager 5

To include a Summary of Findings table in a Cochrane systematic review you can

- **import** a Summary of Findings table which was created in GRADEpro and saved as a *.sof
- **manually create** Summary of Findings table in Review Manager

You can import Summary of Findings table into Review Manager 5 in 2 ways

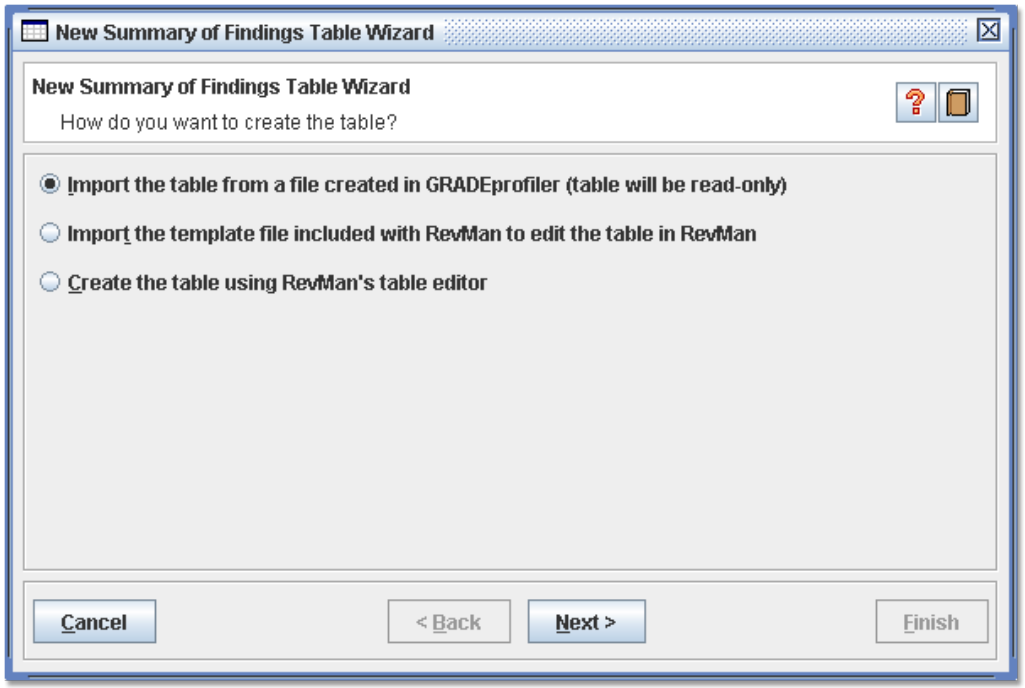
- in RevMan go to **File » Import » Summary of findings table** and choose your *.sof file to import
- in RevMan select **Summary of findings tables** in tree view



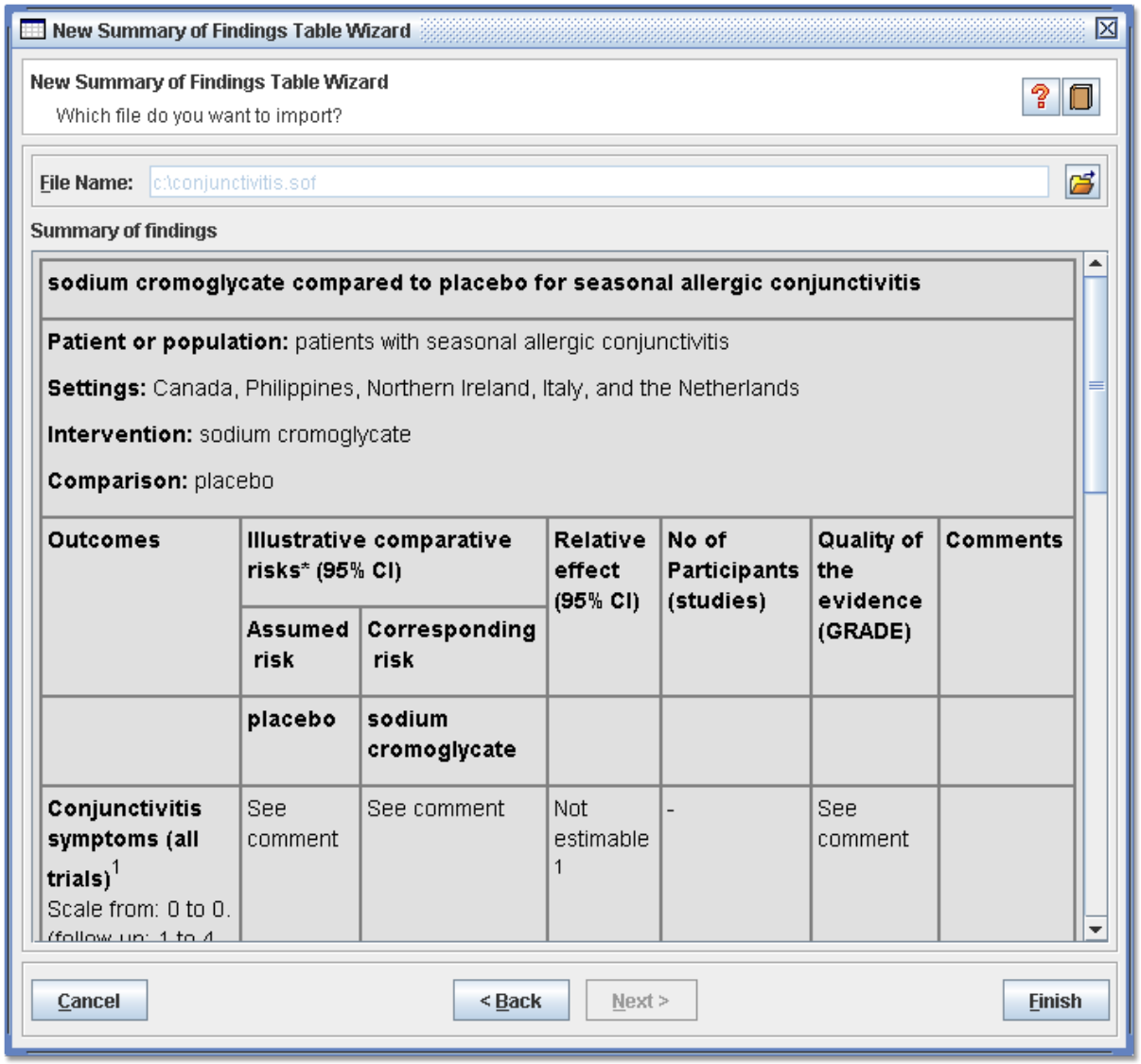
\

- click **Add Summary of Findings Table**

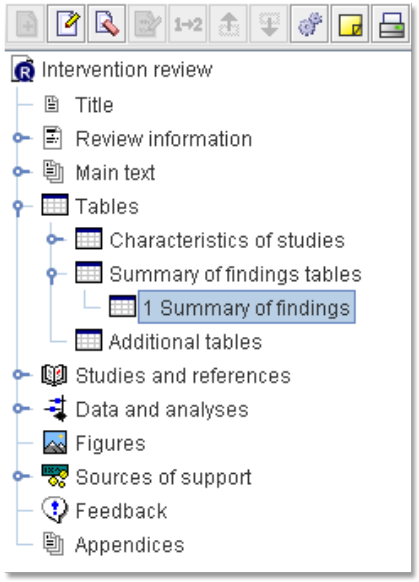
- in a **New Summary of Findings Table Wizard** choose **Import the table from a file created in GRADEprofiler**



- click next and choose your *.sof file in a dialog box that appears
- preview the SoF that will be importedin the wizard



- click Finish (a new table will apear in Review Manager)



Note

The table will appear a little differently than in GRADEprofiler. It is also different from its publishable format. However, the formatting properties have been transferred from GRADEpro to Review Manager 5.0 and **it will be formatted correctly when published** in the Cochrane Library.

7.2.4 Exporting to Microsoft Office Word

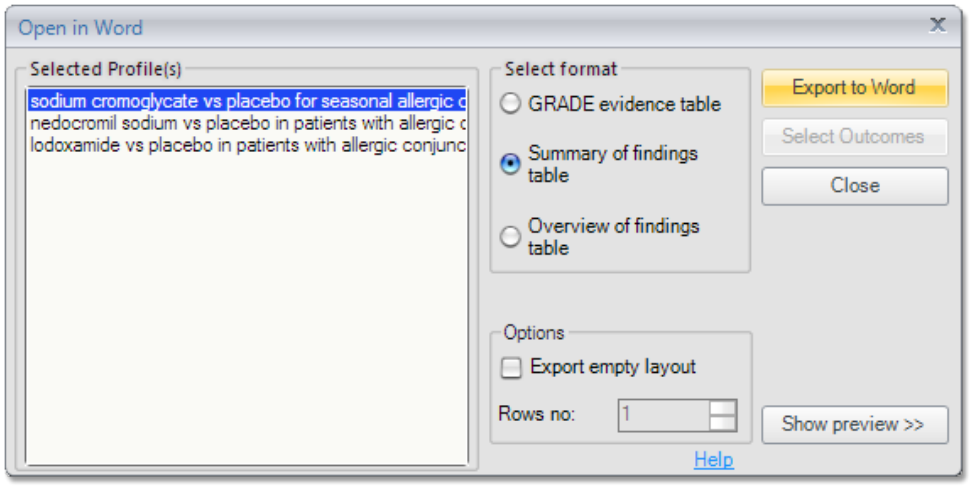
Exporting to Microsoft® Office Word

Note

In order to be able to export Summary of Findings table to Microsoft® Office Word you need to have Microsoft® Office installed on your computer.

To export a table created in GRADEpro into Mictosoft® Office Word and include it in your document you should:

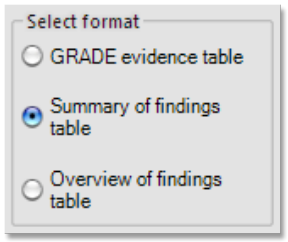
- choose **File » Export As... » Word document**
- in the pop-up window choose the profile(s) you would like to export (to choose many profiles hold CTRL key and click profile names)
- select the format of the table (*e.g.* Summary of Findings table)
- click Export to Word — this will open Microsoft® Office Word with a new document containing your table



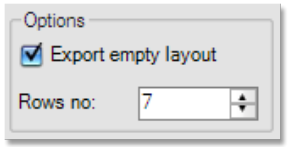
Export Empty Template

On occasion you may wish to save an empty table for later printing or completing in Microsoft® Office Word. That is, **you can produce an empty template** of a GRADE evidence profile, SoF table or Overview of Reviews table from the preview screen.

- **Choose format** of a table you would like to export



- check **Export Empty Layout**



- specify the number of rows by typing in a number or using the arrow buttons
- **Open in Word**.

Note

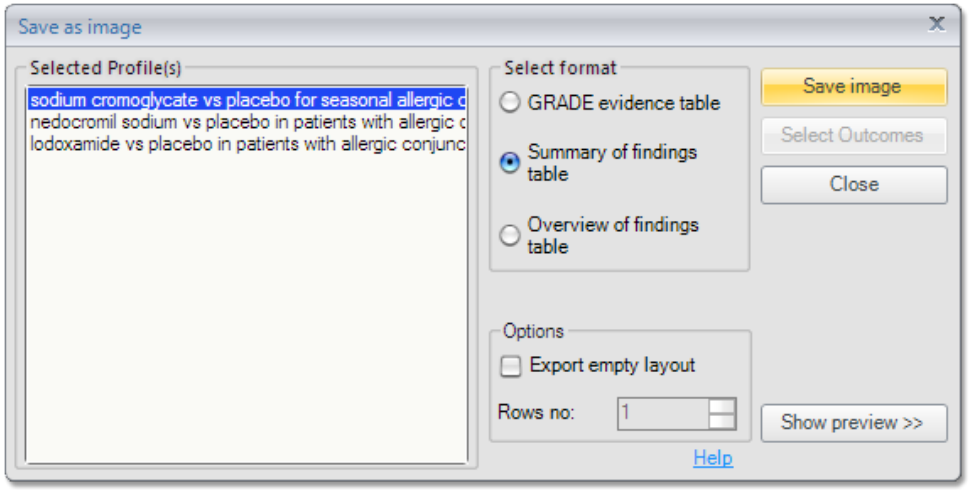
Exporting GRADE evidence profiles to Microsoft® Office Word can be helpful when outcomes are derived from observational studies that include case-control studies as the presentation of results can differ.

7.2.5 Exporting a picture

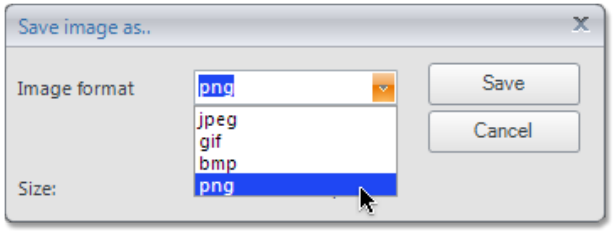
Exporting a Picture

To export a table created in GRADEpro as a picture and include it in your documents you should:

- choose **File » Export As... » Image**
- in the pop-up window choose the profile(s) you would like to export (to choose many profiles hold CTRL key and click profile names)
- select the format of the table (*e.g.* Summary of Findings table)
- click **Save image**



- choose image format — available formats are: JPEG, GIF, BMP and PNG



- **save**

In most situations saving the image of your evidence table in PNG format will be the most reasonable choice.

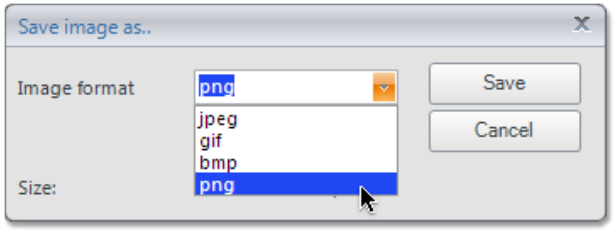
» More information about image formats

7.2.5.1 Image formats

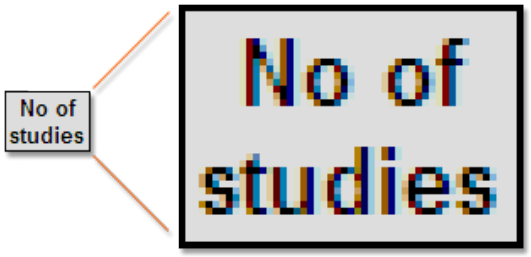
About Image Format

GRADEprofiler lets you export (save) evidence tables as **bitmap images** in 4 different formats

- JPEG
- GIF
- BMP
- PNG

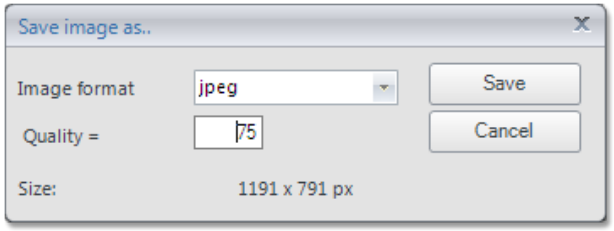


Bitmap (raster) images use a rectangular grid of pixels (dots) to represent images. Bitmap images are resolution-dependent and they lose detail and appear jagged when they are magnified.



JPEG

Joint Photographic Experts Group (JPEG) format is commonly used to display photographs. JPEG format is lossy compression format that reduces file size by selectively discarding data. A higher level of compression results in lower image quality, and a lower level of compression results in better image quality. In most cases Quality = 100 option produces an image indistinguishable from the original. In GRADEprofiler Quality = 75 is set by default. With this empirically determined value it produces files of lowest size maintaining acceptable image quality.



A file size of a representative evidence table might be 624 kB with Quality set to 100, 202 kB with Quality set to 75, and 147 kB with Quality set to 50.

GIF

Graphics Interchange Format (GIF) format is used for lossless compression (however it can save only up to 256 colors or shades of grey). It is used to display images on the web, however, it has recently been replaced by PNG format.

A file size of the same representative evidence table might be 72 kB when saved as GIF.

BMP

BMP format is a standard image format on DOS and Microsoft® Windows compatible computers. Because it is uncompressed it produces very large file sizes.

A file size of the same representative evidence table might be 3.5 MB when saved as BMP.

PNG

Portable Network Graphics (PNG) format is used for lossless compression and is suggested **to display images of evidence tables on the web**.

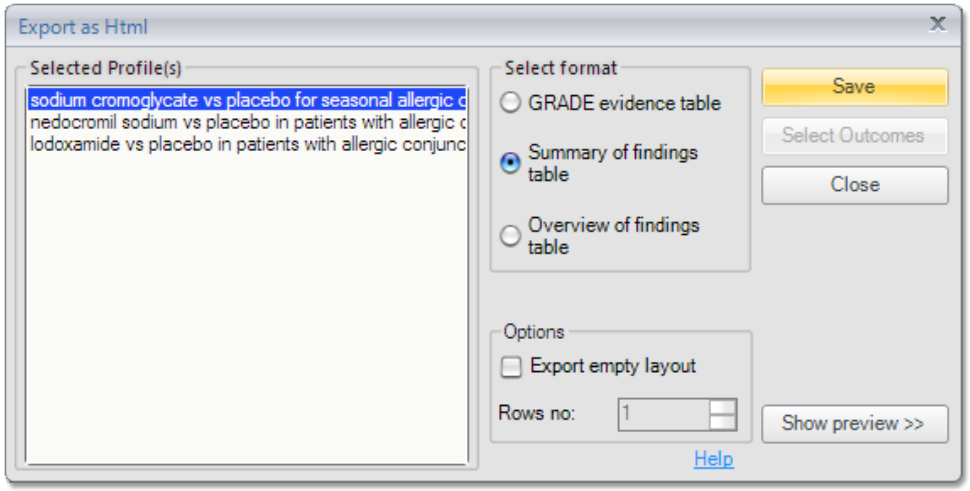
A file size of the same representative evidence table might be 100 kB when saved as PNG.

7.2.6 Exporting HTML

Exporting as HTML

To export a table created in GRADEpro in HTML (HyperText Markup Language) format (*.html or *.htm) and publish it on the web or import to a word processor you should:

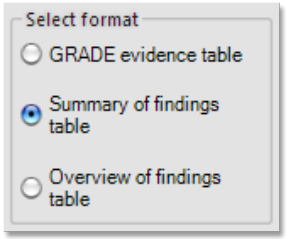
- choose **File » Export As... » HTML** or press **Preview SoF table** on the Toolbar.
- in the pop-up window choose the profile(s) you would like to export (to choose many profiles hold CTRL key and click profile names)
- select the format of the table (*e.g.* Summary of Findings table)
- click **Save**



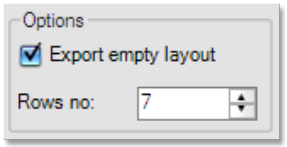
Export Empty Template

On occasion you may wish to save an empty table for later printing or importing it and completing in a word processor. That is, **you can produce an empty template** of a GRADE evidence profile, SoF table or Overview of Reviews table from the preview screen.

- **Choose format** of a table you would like to export



- check **Export Empty Layout**



- specify the number of rows by typing in a number or using the arrow buttons
- **Save**.

7.3 Printing

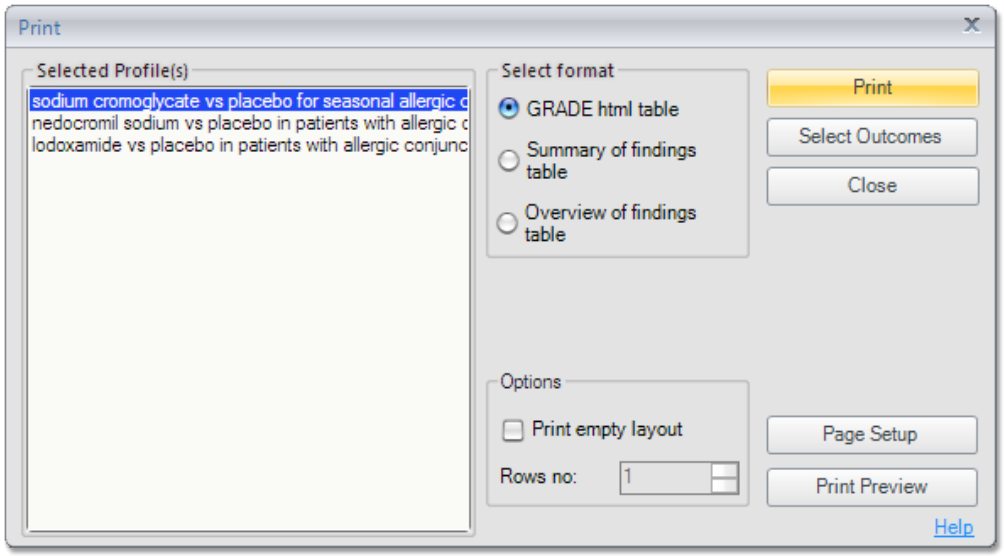
Printing

You can print any table created in GRADEprofiler.

- choose **File » Print**
- in the pop-up window choose the profile(s) you would like to export (to choose many profiles hold CTRL key and click profile names)
- select the format of the table (*e.g.* Summary of Findings table)
- click **Print**

You can preview how your table will look like when printed by clicking **Print Preview**.

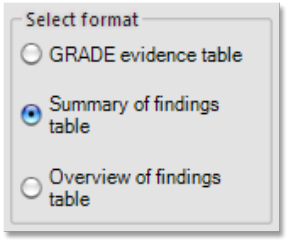
You can set paper size, orientation, margins etc. by clicking **Page Setup**. GRADEprofiler uses standard Microsoft® Internet Explorer printing interface. To obtain best results we strongly recommend that you **update your Internet Explorer to the newest version** (*i.e.* 7.0 or newer).



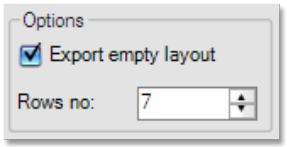
Print Empty Layout

On occasion you may wish to print an empty table for filling on paper.

- **Choose format** of a table you would like to print



- check **Print Empty Layout**



- specify the number of rows by typing in a number or using the arrow buttons
- **Print**.

7.4 Copying, duplicating, and moving

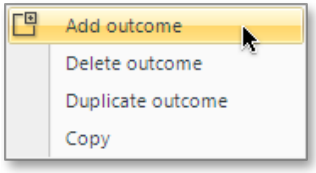
Copying, Duplicating, Moving, and Deleting

In a tree pane you can manage your profile groups, profiles, and outcomes.

By **left-clicking and draging** them you can move an outcome, a profile, or whole profile group.

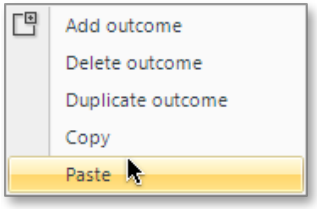
By **right-clicking** them you can copy, duplicate, move, delete, and save whole profile groups or profiles, or single outcomes.

Managing single outcomes

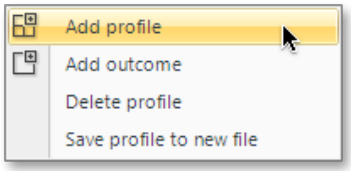


By **right-clicking** an outcome in a tree pane you may

- add a new outcome
- delete the outcome you right-clicked
- duplicate the outcome you right-clicked (this will create a new outcome called "*right-clicked outcome* copy" that is an exact copy of a right-clicked outcome and all its contents, that you may wish to use as a starting point for entering information about another very similar outcome by modifying it insted of entering all the data from the beginning)
- copy the outcome with all the information it contains into a clipboard (this will allow you to paste it into a different profile or even into a different GRADEpro file that is currently open — right-click ahain and select: paste)



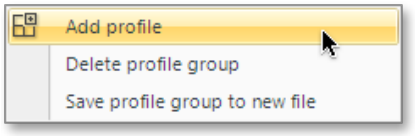
Managing profiles (groups of outcomes)



By **right-clicking** a profile in a tree pane you may

- add a new profile
- add a new outcome at the bottom of this profile branch
- delete the profile you right-clicked
- save the profile you right-clicked to a new GRADEpro file (e.g. to share it with others)

Managing groups of profiles



By **right-clicking** a profile group in a tree pane you may

- add a new profile
- delete the profile group you right-clicked
- save whole profile group you right-clicked to a new GRADEpro file (e.g. to share it with others)

8. Overview of the GRADE Approach

Overview of the GRADE Approach

The GRADE approach is a method of grading the quality of evidence and strength of recommendations in health. It can be used to develop clinical practice guidelines and other health care recommendations. It has been developed by the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) Working Group (www.gradeworkinggroup.org). The working group is a collaboration of methodologists, guideline developers, clinicians and other interested members with the aim of developing and implementing a common, transparent and sensible approach to grading the quality of evidence and strength of recommendations in health care. Membership is open and free.

The GRADE Approach

8.1 The GRADE approach

The GRADE Approach

The GRADE approach is a sequential process for preparing evidence profiles (summaries) and developing evidence-based recommendations.

To achieve a balanced view when formulating recommendations a multidisciplinary panel with broad representation including clinicians, methodologists, generalists, patient representatives

and experienced guideline developers should be assembled and proper group processes for reaching consensus on recommendations should be followed.

When guideline panels or other groups making recommendations decide to follow the GRADE approach, we encourage them to follow the subsequent steps for the recommendations they make:

For guideline panels and others developing healthcare recommendations, the GRADE approach to grading the quality of evidence and strength of recommendations includes the following steps (the shaded area applies to Authors of Systematic Reviews – see note below):

- Ask a specific management question to be answered by a recommendation.
- Identify all important outcomes for every health care question.
- Judge the relative importance of outcomes.
- Summarize all relevant evidence, ideally in evidence profiles.
- Grade the quality of evidence for each outcome.
- Decide on the overall quality of evidence across outcomes.
- Include judgments about the underlying values and preferences related to the management options and outcomes.
- Decide on the balance of desirable and undesirable effects.
- Decide on the balance of net benefits and cost.
- Grade the strength of recommendation.
- Formulate a recommendation.
- Implement and evaluate.

> *For authors of systematic reviews*
>
> Systematic reviews should not include health care recommendations. Therefore, the GRADE approach for grading the quality of evidence terminates after grading the quality of evidence for individual outcomes during the completion of a systematic review. Others, e.g. a guideline panel, will have to complete the subsequent steps.

8.2 Purpose

Purpose of the GRADE System

Clinical Practice Guidelines (CPG) offer recommendations for the management of typical patients. These management decisions involve balancing the desirable and undesirable effects of a given course of action. In order to help clinicians make evidence-based medical decisions, guideline developers often grade the strength of their recommendations.

Prior grading systems had many disadvantages including the lack of separation between quality of evidence and strength of recommendation, the lack of transparency about judgments, and the lack of explicit acknowledgment of values and preferences underlying the recommendations. In addition, the existence of many, often scientifically outdated, grading systems has created confusion among guideline developers as well as users.

The GRADE system was developed to overcome these shortcomings of previous grading systems.

8.3 Separation of quality and strength

Separating Quality of Evidence from Strength of Recommendations

The GRADE approach is based on a sequential assessment of the quality of evidence, followed by judgment about the balance between desirable and undesirable effects, and subsequent decision about the strength of a recommendation.

> **Separating** the judgments regarding the **quality of evidence** from judgments about the **strength of recommendations** is a critical and **defining feature of the GRADE** grading system.

GRADE stresses the necessity to acknowledge the values and preferences underlying the recommendations, and postulates a systematic approach to grading the recommendations that can minimize bias and aid interpretation.

Therefore, unlike many other grading systems, the GRADE system emphasizes that weak recommendations in the face of high quality evidence are common because of factors other than the quality of evidence influencing the strength of a recommendation. For the same reason it allows for **strong recommendations based on the evidence from observational studies**.

Example 1

Weak recommendation based on high quality evidence

Several RCTs compared the use of combination chemotherapy and radiotherapy versus radiotherapy alone in unresectable, locally advanced non-small cell lung cancer (Stage IIIA. The overall quality rating for these trials could be considered high by a guideline panel. Compared with radiotherapy alone, the combination of chemotherapy and radiotherapy reduces the risk for death corresponding to a mean gain in life expectancy of few months, but increases harm and burden related to chemotherapy. Thus, considering the values and preferences patients would place on the small survival benefit in view of the harms and burdens, guideline panels may offer a weak recommendation despite the high quality of the available evidence.

Example 2

Strong recommendation based on low or very low quality evidence

The principle of administering appropriate antibiotics rapidly in the setting of severe infection or sepsis has not been tested against its alternative of no rush of delivering antibiotics in randomized controlled trials. Yet guideline panels who apply the GRADE approach would be very likely to make a strong recommendation for the rapid use of antibiotics in this setting on the basis of available observational studies because the benefits of antibiotic therapy clearly outweigh the downsides in most patients independent of the quality assessment.

8.4 Health care question

About the Health Care Question

A guideline panel should define the scope of each recommendation. Each recommendation should answer a focused and sensible healthcare question. Similarly, authors of each systematic review should formulate focused health care question(s) that the review will answer. A systematic review may answer one or more health care questions, depending on the number of comparisons included in the review.

Every health care management question has four components:

- **P**atients (population)
- **I**nterventions (therapeutic, diagnostic, etc.) under investigation (the experimental intervention or in observational studies this may be exposure)

- **Comparison** (alternative intervention; intervention in the control group)
- **Outcomes** of interest.

Authors of a systematic reviews may address in their review selected outcomes or even one outcome only. However, authors of systematic reviews could make their reviews more useful by looking at a comprehensive range of outcomes.
On the contrary, to make sensible recommendations guideline panels must consider **all outcomes** that are important to patients. **Guideline developers must base outcome choice on what is important, not what was measured.** If evidence is lacking for an important outcome, this should be acknowledged, rather than ignoring the outcome.

Format of health care questions

Defining a health care question includes specifying all outcomes of interest. Recommending to use or not to use a given intervention (therapeutic or diagnostic) one has to consider all outcomes simultaneously, so for the simplicity of the questions covered by the guideline outcomes are not included in the question formats suggested below. However, all outcomes should be listed in the evidence profile.

GRADEpro lets you choose from four different question formats:

- Should **intervention** be used for **health problem**?
- Should **intervention** vs **comparison** be used for **health problem**?
- Should **intervention** be used in **population**?
- Should **intervention** vs **comparison** be used in **population**?

Example

Should manual toothbrushes vs powered toothbrushes be used for dental health?

Should topical nasal steroids be used in children with persistent allergic rhinitis?

Surrogate (substitute) outcomes

Guideline developers should **consider surrogate outcomes only when evidence about patient-important outcomes is lacking**. When this is the case, they should specify the patient-important outcomes and the associated surrogates they must use as substitutes. The necessity to substitute the surrogate may ultimately lead to rating down the quality of the evidence because of the indirectness of the surrogate. Guideline developers should not list the surrogates themselves as their measures of outcome.

Other considerations

A guideline panel may also want to specify the **setting** for which the recommendation is intended (*e.g.* outpatient vs. inpatient).

When making recommendations, one should define **questions that lead to action**. When evaluating etiological information it should be made clear whether the intended action is the removal of a potential etiological agent. This is because action follows only from knowledge that modifying etiological factors influences outcomes. These actions can be based on indirect evidence from exposure assessment and the resulting indirect judgment that removal of the exposure would reduce any harmful effects.

Some frequent errors in formulating the health care question include:

- failure to include all patient-important outcomes (*e.g.* adverse effects or toxicity)
- inclusion of surrogate outcomes, in particular when information on patient-important outcomes is not available, *e.g.* using exercise capacity rather than quality of life, or bone mineral density rather than fracture rate
- failure to fully consider all relevant alternatives (this may be particularly problematic when guidelines target a global audience)

8.5 Importance of outcomes

About the Importance of the Outcomes

GRADE specifies three categories of outcomes according to their **importance**. Guideline developers must, and authors of systematic reviews are strongly encouraged to specify all potential patient-important outcomes as the first step in their endeavour. Those creating evidence profiles will also classify outcomes as:

- critical
- important but not critical
- of limited importance.

The first two classes of outcomes will bear on guideline recommendations; the third may or may not. Ranking outcomes by their relative importance can help to focus attention on those outcomes that are considered most important, and help to resolve or clarify disagreements.

Different audiences are likely to have different perspectives on the importance of outcomes

The importance of outcomes is likely to vary within and across cultures or when considered from the perspective of patients, clinicians or policy-makers. It is essential to take cultural diversity into account when deciding on relative importance of outcomes, particularly when developing recommendations for an international audience. Guideline panels should also decide what perspective they are taking. Our view is that this should usually be the patients' perspective. Guideline panels may also choose to take the perspective of society as a whole (*e.g.* a guideline panel developing recommendations about pharmacological management of bacterial sinusitis may take the patient perspective when considering health outcomes, but also a society perspective when considering antimicrobial resistance to specific drugs).

Rating	Importance
9	critical
8	
7	
6	important
5	
4	
3	not important
2	
1	

Guideline developers should first consider whether particular desirable or undesirable consequences of a therapy are **important** to the decision regarding the optimal management strategy, or whether they are **of limited importance**. If the guideline panel thinks that a particular outcome is important, then it should consider whether the outcome is **critical** to the decision, or only important, but not critical. GRADE suggests using a 9-point scale to rate importance of the outcomes.

To facilitate ranking of outcomes according to their importance guideline developers as well as authors of systematic reviews may choose to rate outcomes numerically on a **1 to 9 scale** (7 to 9 – critical; 4 to 6 – important; 1 to 3 – of limited importance) to distinguish between importance categories.

One should aim to decide which outcomes are important during protocol development and **before one undertakes a systematic review or guideline project**. However, rating importance of an outcome prior to evidence review is preliminary – when evidence becomes available a reassessment of importance is necessary. Guideline panels should be aware of the possibility that in some instances the importance of an outcome (e.g. a serious adverse effect) may only become known after the protocol was written, evidence was reviewed or the analyses were carried out, and should take appropriate actions to include these in the evidence tables. **Outcomes that are critical to decision making should be included in an evidence table whether or not information about them is available.**

- Only outcomes considered **important** (rated 4—9) or **critical** (rated 7—9) should be included in the **evidence profile**.
- Only outcomes considered **critical** (rated 7—9) are the primary factors influencing a recommendation and should be used to determine the **overall quality of evidence** supporting this recommendation.

8.6 Summarizing the evidence

About Summary of the Relevant Evidence

A guideline panel should base its recommendation on the **best available evidence** related to the health care question.

A guideline panel can use already existing high quality **systematic reviews** or conduct its own systematic review depending on the specific circumstances such as availability of high quality systematic reviews and resources, but GRADE recommends that systematic reviews should form the basis for making healthcare recommendations. One should seek evidence relating to **all patient-important outcomes** and for the **values** patients place on these outcomes as well as related management options.

The evidence collected from systematic reviews is used to produce evidence tables including the assessment of quality of evidence and a summary of findings (the effect size in the intervention and comparison groups, and the magnitude relative and the absolute effects). Thus, a GRADE evidence profile including the assessment of quality of evidence and a summary of findings is a transparent summary of evidence on which those making recommendations can base their judgments.

8.7 Quality of evidence

About Quality of Evidence

Quality of evidence

- Definition
- Grades of quality of evidence
- Factors determining the quality of evidence for single outcome
- Deciding on the overall quality of evidence across outcomes

GRADE provides a specific **definition** of the quality of evidence, that is different in the context of making recommendations and in the context of summarising the findings of a systematic review.

For guideline panels

The quality of evidence reflects the extent to which our **confidence in an estimate of the effect** is adequate to **support a particular recommendation**.
Guideline panels must make judgments about the quality of evidence relative to the specific context for which they are using the evidence.
The GRADE approach involves separate grading of quality of evidence for each patient-important outcome followed by determining an overall quality of evidence across outcomes.

For authors of systematic reviews

The quality of evidence reflects the extent to which we are **confident that an estimate of the effect is correct**.
Because systematic reviews do not – or at least should not – make recommendations, they require a different definition. Authors of systematic reviews grade

quality of a body of evidence separately for each patient-important outcome.

To achieve transparency and implicity, the GRADE system classifies the quality of evidence in one of **four grades**:

Grade	Definition
High	Further research is very unlikely to change our confidence in the estimate of effect.
Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low	Any estimate of effect is very uncertain.

8.7.1 Factors determining the quality of evidence

Factors Determining the Quality of Evidence

Initial rating

Study design

Factors that can reduce the quality of the evidence

Factor	Consequence
Limitations in study design or execution (risk of bias)	↓ 1 or 2 levels
Inconsistency of results	↓ 1 or 2 levels
Indirectness of evidence	↓ 1 or 2 levels
Imprecision	↓ 1 or 2 levels
Publication bias	↓ 1 or 2 levels

Factors that can increase the quality of the evidence

Factor	Consequence
Large magnitude of effect	↑ 1 or 2 levels
All plausible confounding would reduce the demonstrated effect or increase the effect if no effect was observed	↑ 1 level
Dose-response gradient	↑ 1 level

While factors influencing the quality of evidence are **additive** — such that the reduction or increase in each individual factor is added together with the other factors to reduce or increase the quality of evidence for an outcome — grading the quality of evidence involves judgements which are not exclusive. Therefore, GRADE is not a quantitative system for grading the quality of evidence. Each factor for downgrading or upgrading reflects **not discrete categories but a continuum** within each category and among the categories. When the body of evidence is intermediate with respect to a particular factor, the decision about whether a study falls above or below the threshold for up- or downgrading the quality (by one or more factors) depends on judgment.
For example, if there was some uncertainty about the three factors: study limitations, inconsistency, and imprecision, but not serious enough to downgrade each of them, one could reasonably make the case for dowgrading, or for not doing so. A reviewer might in each category give the studies the benefit of the doubt and would interpret the evidence as high quality. Another reviewer, deciding to dowgrade the evidence by one level, would judge the evidence as moderate quality. Reviewers should grade the quality of the evidence by considering both the individual factors in the context of other judgments they made about the quality of evidence for the same outcome.

In such a case, you should pick one or two categories of limitations which you would offer as reasons for downgrading and explain your choice in the footnote. You should also provide a footnote next to the other factor, you decided not to downgrade, explaining that there was some uncertainty, but you already downgraded for the other factor and further lowering the quality of evidence for this outcome would seem inappropriate. GRADE strongly encourages review and guideline authors to be **explicit and transparent** when they find themselves in these situations by **acknowledging borderline decisions**.

Despite the limitations of breaking continua into categories, treating each criterion for rating quality up or down as discrete categories enhances transparency. Indeed, the **great merit of GRADE** is not that it ensures reproducible judgments but that it **requires explicit judgment** that is made **transparent to users**.

Note

Observational studies that have been **downgraded to very low quality** for any reason **should not be upgraded**.

click for larger image

You may follow the GRADE diagram to facilitate grading of the quality of evidence.

You may also download and print the diagram. [You will need Acrobat Reader in order to view this document]

8.7.1.1 Study design

About Study Design

Study design is critical to judgments about the quality of evidence.

For recommendations regarding management strategies – as opposed to establishing prognosis or the accuracy of diagnostic tests – **randomized trials** provide, in general, far stronger evidence than observational studies, and rigorous **observational studies** provide stronger evidence than **uncontrolled case series**.

In the GRADE approach to quality of evidence

- **randomized trials** without important limitations provide high quality evidence
- **observational studies** without special strengths or important limitations provide low quality evidence

Limitations or special strengths can, however, **modify** the quality of the evidence of both randomized trials and observational studies.

Note

Non-randomised experimental trials (quasi-RCT) without important limitations also provide high quality evidence, but will automatically be downgraded for limitations in design (risk of bias) – such as lack of concealment of allocation and tie with a provider (*e.g.* chart number).

Case series and **case reports** are observational studies that investigate only patients exposed to the intervention. Source of control group results is implicit or unclear, thus, they will usually warrant downgrading from low to very low quality evidence.

Expert opinion is not a category of quality of evidence. Expert opinion represents an interpretation of evidence in the context of experts' experiences and knowledge. Experts may have opinion about evidence that may be based on interpretation of studies ranging from uncontrolled case series (e.g. observations in expert's own practice) to randomised trials and systematic reviews known to the expert. It is important to describe what type of evidence (whether published or unpublished) is being used as the basis for interpretation.

» Compare diagram.

8.7.1.2 Study limitations (risk of bias)

About Study Limitations (Risk of Bias)

Limitations in the study design and implementation may **bias** the estimates of the treatment effect. Our confidence in the estimate of the effect and in the following recommendation decreases if studies suffer from major limitations.
The more serious limitations are, the more likely it is that the quality of evidence will be downgraded. Numerous tools exist to evaluate the risk of bias in randomized trials and observational studies. This handbook describes the key criteria used in the GRADE approach.

Our confidence in an estimate of effect decreases if studies suffer from major limitations that are likely to result in a biased assessment of the intervention effect. For randomized trials, the following limitations are likely to result in biased results.

Limitations of randomized controlled trials

LIMITATION	EXPLANATION
Lack of allocation concealment	Those enrolling patients are aware of the group to which the next enrolled patient will be allocated (major problem in "pseudo" or "quasi" randomized trials with allocation by day of week, birth date, chart number etc.)
Lack of blinding	Patient, caregivers, those recording outcomes, those adjudicating outcomes, or data analysts are aware of the arm to which patients are allocated
Incomplete accounting of patients and outcome events	Loss to follow-up and failure to adhere to the intention to treat principle when indicated
Selective outcome reporting	Reporting of some outcomes and not others on the basis of the results
Other limitations	For example: ■ stopping early for benefit observed in randomized trials, in particular in the absence of adequate stopping rules ■ use of unvalidated patient-reported outcomes

- carry-over effects in cross-over trials
- recruitment bias in cluster-randomized trials

Systematic reviews of tools to assess the methodological quality of non-randomized studies have identified over 200 checklists and instruments. We summarize here key criteria for observational studies that reflect the contents of these checklists.

Limitations of observational studies

LIMITATION	EXPLANATION
Failure to develop and apply appropriate eligibility criteria (inclusion of control population)	■ under- or over-matching in case-control studies ■ selection of exposed and unexposed in cohort studies from different populations
Flawed measurement of both exposure and outcome	■ differences in measurement of exposure (e.g. recall bias in case- control studies) ■ differential surveillance for outcome in exposed and unexposed in cohort studies
Failure to adequately control confounding	■ failure of accurate measurement of all known prognostic factors ■ failure to match for prognostic factors and/or adjustment in statistical analysis
Incomplete or inadequately short follow-up	

Depending on the context and study type, there can be additional limitations than those listed above. Guideline panels and authors of systematic reviews should consider all possible limitations.

Guideline panels or authors of systematic reviews should consider the extent to which study limitations may bias the results. If the limitations are serious they may downgrade the quality rating by one or even two levels.

» Compare diagram.

For authors of systematic reviews

Systematic reviewers working within the context of Cochrane Systematic Reviews, can use the following guidance to assess study limitations (risk of bias) in Cochrane Reviews. Chapter 8 of the Cochrane Handbook provides a detailed discussion of study-level assessments of risk of bias in the context of a Cochrane review, and proposes an approach to assessing the risk of bias for an outcome across studies as 'low risk of bias', 'unclear risk of bias' and 'high risk of bias' (Cochrane Handbook Chapter 8, Section 8.7). These assessments should feed directly into the assessment of study limitations. In particular, 'low risk of bias' would indicate 'no limitation'; 'unclear risk of bias' would indicate either 'no limitation' or 'serious limitation'; and 'high risk of bias' would indicate either 'serious limitation' or 'very serious limitation' in the GRADE approach. Cochrane systematic review authors must use their judgment to decide between alternative categories, depending on the likely magnitude of the potential biases.

Every study addressing a particular outcome will differ, to some degree, in the risk of bias. Review authors must make an overall judgment on whether the quality of evidence for an outcome warrants downgrading on the basis of study limitations. The assessment of study limitations should apply to the studies contributing to the results in the 'Summary of findings' table, rather than to all studies that could potentially be included in the analysis.

Going from assessments of study limitations (risk of bias) to judgements about study limitations for outcomes

RISK	GRADE ASSESSMENT OF

OF BIAS	ACROSS STUDIES	INTERPRETATION	CONSIDERATIONS	STUDY LIMITATIONS
Low	Most information is from studies at low risk of bias.	Plausible bias unlikely to seriously alter the results.	No apparent limitations.	No serious limitations, do not downgrade.
Unclear	Most information is from studies at low or unclear risk of bias.	Plausible bias that raises some doubt about the results.	Potential limitations are unlikely to lower confidence in the estimate of effect.	No serious limitations, do not downgrade.
			Potential limitations are likely to lower confidence in the estimate of effect.	Serious limitations, downgrade one level.
High	The proportion of information from studies at high risk of bias is sufficient to affect the interpretation of results.	Plausible bias that seriously weakens confidence in the results.	Crucial limitation for one criterion, or some limitations for multiple criteria, sufficient to lower confidence in the estimate of effect.	Serious limitations, downgrade one level.
			Crucial limitation for one or more criteria sufficient to substantially lower confidence in the estimate of effect.	Very serious limitations, downgrade two levels.

EXAMPLE 1

Blinding

RCTs of the effects of Intervention A on acute spinal injury measured both all-cause mortality and, based on a detailed physical examination, motor function. Blinding of outcome assessors is less important for the assessment of all-cause mortality, but crucial for motor function. If the studies are not blinded, then the quality of the evidence for mortality may not be downgraded but may be downgraded for motor function.

EXAMPLE 2

Overall

Three RCTs of the effects of surgery on patients with lumbar disc prolapse measured symptoms after 1 year or longer. The RCTs suffered from inadequate concealment of allocation, and unblinded assessment of outcome by potentially biased raters (surgeons) using unvalidated rating instrument. The benefit of surgery is uncertain. The quality of the evidence was downgraded by two levels due to these study limitations quality.

8.7.1.3 Inconsistency of results

About Unexplained Heterogeneity or Inconsistency of Results

Inconsistency refers to an **unexplained heterogeneity** of results.

Widely differing estimates of the treatment effect (i.e. heterogeneity or variability in results) across studies suggest true differences in underlying treatment effect. When heterogeneity exists, but investigators fail to identify a plausible explanation, the quality of evidence should be downgraded by one or two levels, depending on the magnitude of the inconsistency in the results.

Inconsistency may arise from differences in:

- populations (e.g. drugs may have larger relative effects in sicker populations)
- interventions (e.g. larger effects with higher drug doses)
- outcomes (e.g. diminishing treatment effect with time).

Guideline panels or authors of systematic reviews should also consider the extent to which they are uncertain about the underlying effect due to the inconsistency in results and they may downgrade the quality rating by one or even two levels.

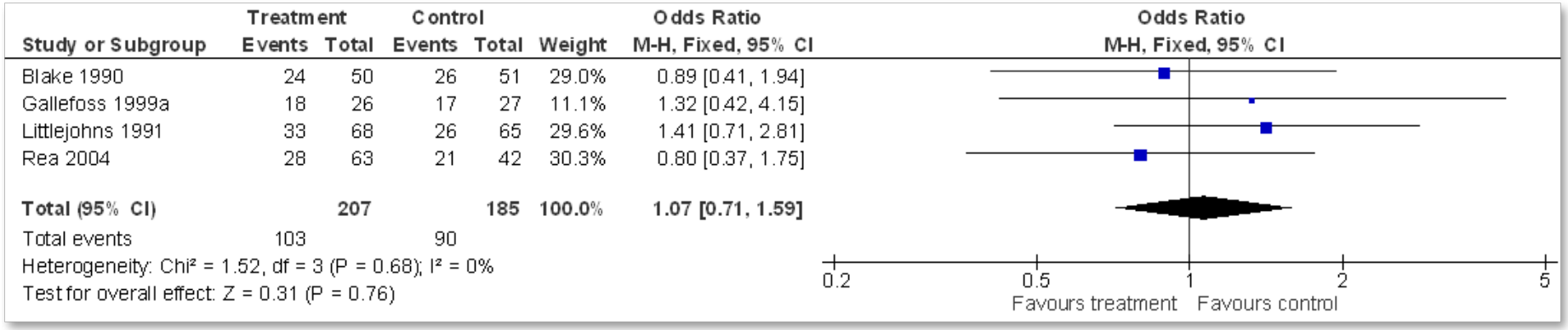
If inconsistency can be explained, separate quality assessments for the stratification that explains the observed heterogeneity should be completed (that includes the possibility that guideline panels make separate recommendations based on the identified explanatory factors, *i.e.* population and intervention). For instance, separate evidence tables would be completed for symptomatic patients with high degree stenosis of the carotid artery in which the intervention is endarterectomy (beneficial in the hands of the right surgeons), and another (if they considered it worth it) for asymptomatic patients with moderate degree stenosis in which surgery is not beneficial. When heterogeneity exists and affects the interpretation of results, but authors fail to identify a plausible explanation, the quality of evidence decreases.

» Compare diagram.

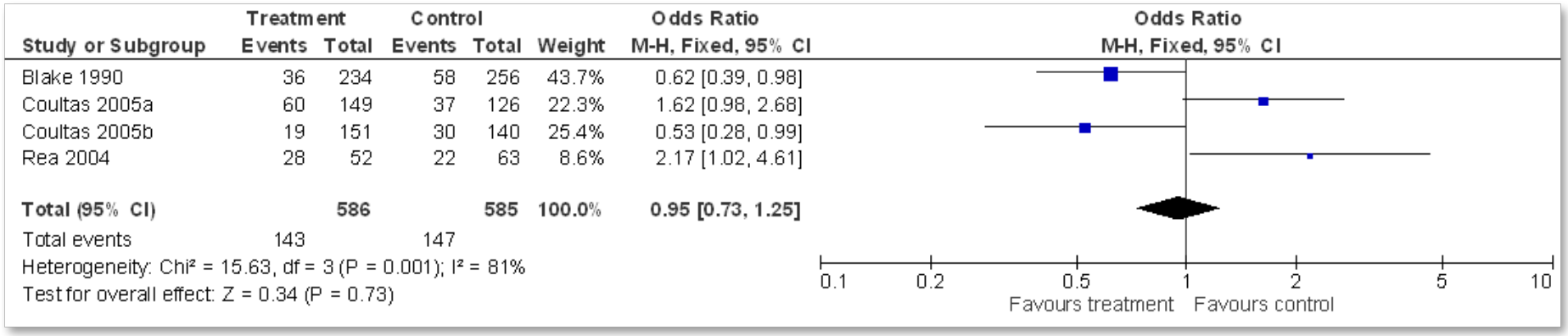
Example 1

Direction of effect is not a criterion for inconsistency

Consider the forest plot below, with four studies, two on either side of the line of no effect, where the confidence intervals overlap, the p value for heterogeneity is greater than 0.05, and I^2 is 0. The quality of the evidence would not be downgraded for inconsistency based on the fact that the point estimates are compatible with benefit and harm.



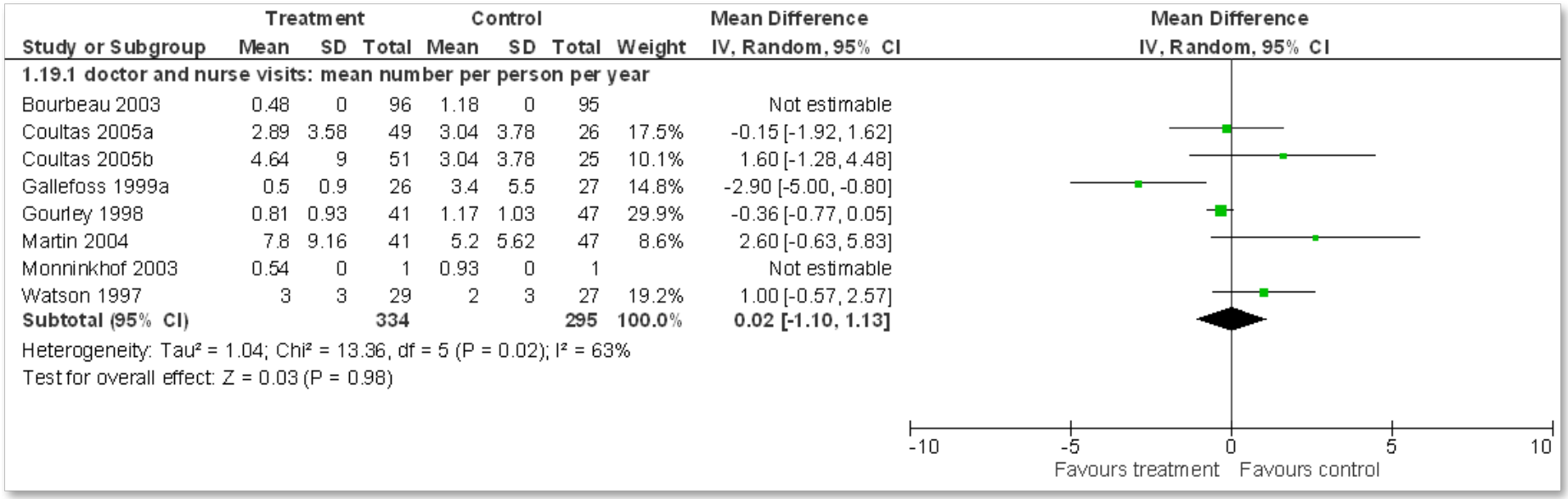
Consider the forest plot below with four studies, where the confidence intervals do not overlap, the p value for heterogeneity is less than 0.05, and I^2 is 81%. Two studies suggest benefit and the other two harm. The quality of the evidence would be downgraded for inconsistency. However, they could be downgraded for inconsistency whether or not they would fall on either side of the null effect because of the large differences in the effects between studies.



Example 2

Unexplained heterogeneity may be a reason to downgrade

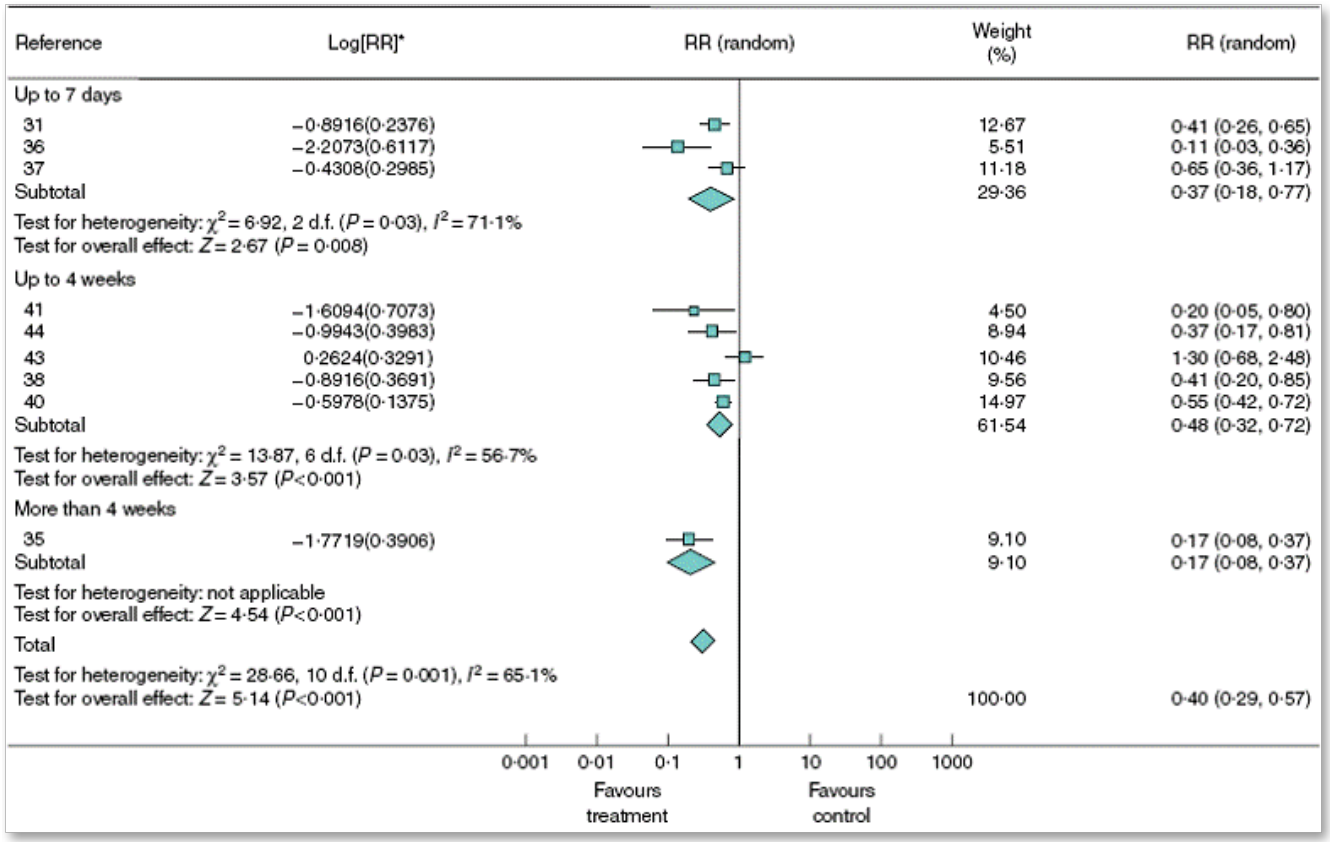
Consider the forest plot below, with 8 studies, confidence intervals with minimal overlap, the p value for heterogeneity less than 0.05 and I2 62.6%. Heterogeneity was not explained by study design, differences in population (*e.g.* final values or change score in control groups) or interventions, or length of follow-up. The quality of the evidence was downgraded from high to moderate quality.



Example 3

Heterogeneity may or may not be a reason to downgrade when results are variable but the differences are between small and large beneficial effects

Consider the forest plot below in which the p value for heterogeneity is lower than 0.05, I^2 is 65.1%. Heterogeneity could not be explained by study design, differences in population/interventions/outcomes, or length of follow-up. All studies, except 1 favours treatment. Since there is appreciable effect, the quality of the evidence may or may not be downgraded for inconsistency. The decision for downgrading would be influenced by factors such as whether the intervention is associated with benefit in all other outcomes or whether the uncertainty about the magnitude of benefit (or harm) of the outcome showing heterogeneity would influence the overall judgment about net benefit or harm (across all outcomes).



8.7.1.4 Indirectness of evidence

About Indirectness of Evidence

There are two types of indirectness.

1. **Indirect comparison** – occurs when a comparisons of intervention A versus B is not available, but A was compared with C and B was compared with C. Such studies allow indirect comparisons of the magnitude of effect of A versus B. Such evidence is of lower quality than head-to-head comparisons of A and B would provide.

2. **Indirect population, intervention, comparator, or outcome** – the question being addressed by the guideline panel or by the authors of a systematic review is different from the available evidence regarding the population, intervention, comparator, or an outcome.

Those making recommendations or authors of systematic reviews should consider the extent to which they are uncertain about the applicability of the evidence to their relevant question and downgrade the quality rating by one or even two levels.

» Compare diagram.

Examples of indirect evidence

Indirect	Question of interest	Source of indirectness
comparison	Relative effectiveness of alendronate and risedronate in osteoporosis.	Randomized trials compared alendronate to placebo and risedronate to placebo, but trials comparing alendronate to risedronate are unavailable.
population	Oseltamivir for prophylaxis of avian flu caused by influenza A (H5N1) virus.	Randomized trials of oseltamivir are available for patients with seasonal influenza, but not for avian influenza.
intervention	Sigmoidoscopic screening for prevention of colon cancer mortality.	Randomized trials of fecal occult blood screening provide indirect evidence bearing on the potential effectiveness of sigmoidoscopy.
comparator	Choice of medication for schizophrenia.	A series of trials comparing newer generation neuroleptic agents to fixed doses of 20 mg of haloperidol provide indirect evidence of how the newer agents would compare to the lower, flexible doses of haloperidol clinicians typically use.
outcome	Rosiglitazone for prevention of diabetic complications in patients at high risk of developing diabetes.	Randomized trials show delay in the development of biochemical diabetes with rosiglitazone, but were underpowered to address diabetic complications.

For authors of systematic reviews

Directness is judged by the users of the evidence tables, depending on the target population, intervention, and outcomes of interest. However, authors of systematic reviews should answer a health care question they asked and thus they may rate the directness of evidence they found. The more clearly and explicitly the health care question was formulated the easier it will be for the users to understand authors' judgment.

Examples

The quality of the evidence may be downgraded when substitute measurements or surrogate endpoints are measured instead of patient-important outcomes.

Here are examples of surrogate measures and coinciding patient-important outcomes common in current clinical investigation:

Condition	Patient-important outcome(s)	Surrogate outcome(s)
Diabetes	Diabetic symptoms, admission, complications (cardiovascular, eye, renal, neuropathic etc.)	Glucose, HbA_{1C}
Dementia	Patient function, behaviour, caregiver burden	Cognitive function
Osteoporosis	Fractures	Bone density
ARDS	Mortality	Oxygenation
End-stage renal disease	Quality of life, mortality	Hemoglobin
Venous thrombosis	Symptomatic venous thrombosis	Asymptomatic venous thrombosis
Chronic respiratory disease	Quality of life, exacerbations, mortality	Pulmonary function, exercise capacity
Cardiovascular disease/risk	Vascular events, mortality	Serum lipids

---

8.7.1.5 Imprecision

About Imprecision (Random Error)

Results are **imprecise** when studies include relatively **few patients and few events** and thus have **wide confidence intervals** around the estimate of the effect. In this case guideline panel will judge the quality of the evidence lower than it otherwise would because of resulting **uncertainty in the results**.

Because GRADE defines the quality of evidence differently for systematic reviews and for guidelines, the criteria for downgrading for imprecision differ in that guideline panels need to consider the context of a recommendation and other outcomes whereas judgments about specific outcomes in a systematic review are free of that context. GRADE approach suggests more specific guidance for determination of imprecision:

- for authors of systematic reviews
- for guideline panels

Copyright © 2004–2009 GRADE Working Group

8.7.1.5.1 Imprecision in systematic reviews

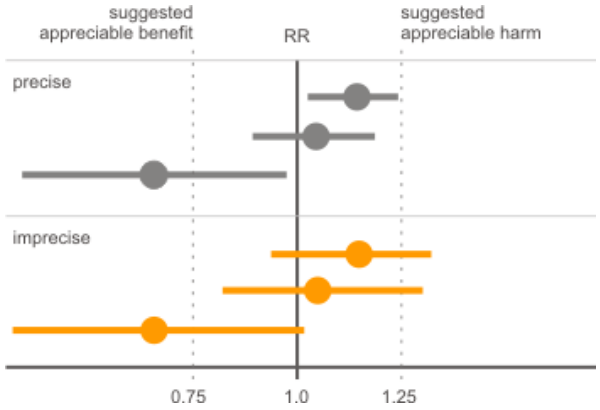
About Imprecision (Random Error) for Authors of Systematic Reviews

For systematic reviews, quality of evidence refers to one's **confidence in the estimates of effect**. In systematic reviews **each outcome is considered separately**.

For dichotomous outcomes

You should consider downgrading the quality of evidence because of imprecision for **either** of the following two reasons:

1. total (cumulative) sample size is lower than the calculated optimal information size (OIS)
   and/or
   total number of events is less than 300 (a threshold rule-of-thumb value)
   (based on: Mueller et al. Ann Intern Med. 2007;146:878-881)
2. 95% confidence interval (or alternative estimate of precision) around the pooled or best estimate of effect includes both 1) no effect and 2) appreciable benefit or appreciable harm. GRADE suggests that the threshold for "appreciable benefit" or "appreciable harm" that should be considered for downgrading is a relative risk reduction (RRR) or relative risk increase (RRI) greater than 25%.



Exception
When **event rates are very low**, 95% confidence intervals around relative effects can be very wide, but 95% confidence intervals around **absolute effects may be narrow**. Under such circumstances you may not downgrade the quality of evidence for imprecision.

For continuous outcomes

We suggest downgrading the quality of evidence when:

1. total (cumulative) sample size is lower than the calculated optimal information size (OIS)
   and/or
   total population size is less than 400 (a threshold rule-of-thumb value; using the usual a and ß, and an effect size of 0.2 SD, representing a small effect)
2. 95% confidence interval includes no effect and the upper or lower confidence limit crosses the minimal important difference (MID), either for benefit of harm (*Note:* if the MID is not known or the use of different outcomes measures required calculation of an effect size (ES), we suggest downgrading if the upper or lower confidence limit crosses an effect size of 0.5 in either direction).

Whether you will rate down for imprecision is **critically dependent on the choice of the difference** (Δ) you wish to detect and the resulting sample size required (OIS). Again, the merit of the GRADE approach is not that it ensures agreement between reasonable individuals, but the explicitness of the judgements being made.

8.7.1.5.2 Imprecision in guidelines

About Imprecision (Random Error) for Guideline Panels

For guidelines, quality refers to the extent to which our **confidence in the estimate of an effect is adequate to support a particular decision**. In guidelines **all outcomes are considered together**, with attention to whether they are critical or important but not critical.

For dichotomous outcomes

We suggest downgrading the quality of evidence because of imprecision for the same criteria described for systematic reviews.

If the evidence meets these criteria, however, guideline developers must in addition consider whether the evidence is adequate to support their recommendation.
We suggest that guideline developers **consider downgrading** the quality of evidence because of imprecision in the following situations:

1. when the recommendation is **in favour** of an intervention and
    a. the 95% confidence interval (or alternative estimate of precision) around the pooled or best estimate of effect **includes no effect** and the upper confidence limit includes an effect that, if it were real, would represent a **benefit that would outweigh the downsides**
    b. the 95% confidence interval (or alternative estimate of precision) around the pooled or best estimate of effect **excludes no effect**, but the lower confidence limit crosses a threshold below which, given the downsides of the intervention, **one would not recommend the intervention**
2. when the recommendation is **against** an intervention and the 95% confidence interval (or alternative estimate of precision) around the pooled or best estimate of effect
    a. the 95% confidence interval (or alternative estimate of precision) around the pooled or best estimate of effect **includes no effect** and the lower confidence limit includes an effect that, if it were real, would represent a harm that, given the benefits, **would still be unacceptable**
    b. the 95% confidence interval (or alternative estimate of precision) around the pooled or best estimate of effect **excludes no effect** but the upper confidence limit crosses a threshold above which, given the benefits of an intervention, **one would recommend the intervention**.

Exception

When event rates are very low, 95% confidence intervals around relative effects can be very wide, but 95% confidence intervals around absolute effects may be narrow. Therefore, if there are very few or no events and the number of participants is large, judgment about precision may be based on the absolute effect. Under such circumstances one may not downgrade the quality of evidence for imprecision.

For continuous outcomes

In the context of a guideline, we suggest using the same criteria described for continuous outcomes in the context of a systematic review, but the decision to downgrade the quality of evidence for imprecision also requires consideration of the full context and the other outcomes as described above.

8.7.1.6 Publication bias

About Publication Bias

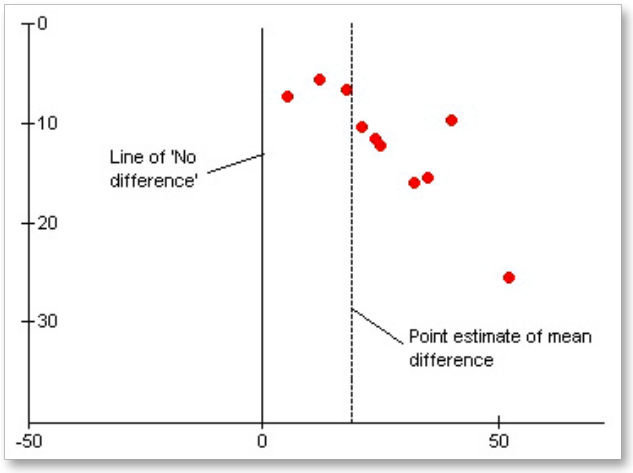
Publication bias is a systematic underestimate or an overestimate of the underlying beneficial or harmful effect due to the **selective publication of studies**.

Publication bias arises when investigators fail to report studies they have undertaken (typically those that show no effect). Methods to detect the possibility of publication bias in systematic reviews exist,although authors of the reviews and guideline panels must often guess about the likelihood of publication bias. A prototypical situation that should elicit suspicion of publication bias occurs when published evidence is limited to a small number of trials, all of which are showing benefits of the studied intervention.

Guideline panels or authors of systematic reviews should consider the extent to which they are uncertain about the magnitude of the effect due to selective publication of studies and they may downgrade the quality of evidence by one or even two levels.

Example

A number of small trials from a systematic review of oxygen therapy in patients with chronic obstructive pulmonary disease showed that the intervention improved exercise capacity, but evaluation of the data suggested publication bias.
Bradley JM, O'Neill B. Short-term ambulatory oxygen for chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews 2005, Issue 4. Art. No.: CD004356. DOI: 10.1002/14651858.CD004356.pub3.



Funnel plot of exercise distance with distance on the x-axis and variance on the y-axis. The blue dots represent the mean differences of individual trial estimates and the dotted line the point estimate of the mean effect indicating benefit from oxygen treatment. The distribution of these dots to the right of the dotted line suggests that there may be the equivalent number of 'negative' trials that have not been included in this analysis.

» Compare diagram.

8.7.1.7 Large or very large effect

About Large Magnitude of the Effect

When methodologically strong observational studies yield large or very large and consistent estimates of the magnitude of a treatment or exposure effect, we may be confident about the results. In those situations, the weak study design is unlikely to explain all of the apparent benefit or harm, even though observational studies are likely to provide an overestimate of the true effect.

The larger the magnitude of effect, the stronger becomes the evidence.

Magnitude of effect	Effect measure	Quality of evidence
large	RR >2 or <0.5 (based on consistent evidence from at least 2 studies, with no plausible confounders)	upgrade 1 level
very large	RR >5 or <0.2 (based on direct evidence with no major threats to validity)	upgrade 2 levels

(RR - relative risk)

Only studies with no threats to validity (not downgraded for any reason) can be upgraded.

» Compare diagram.

Examples

1. A meta-analysis of observational studies showed that bicycle helmets reduce the risk of head injuries in cyclists. This large effect (OR: 0.31, 95%CI: 0.26 to 0.37) suggests a rating of moderate quality evidence.

2. A meta-analysis of observational studies showed that warfarin prophylaxis reduces the risk of thromboembolism in patients with cardiac valve replacement. This very large effect (RR: 0.17, 95%CI: 0.13 to 0.24) suggests a rating of high quality evidence.

8.7.1.8 Plausible biases underestimate true effect

About the Effect of all Plausible Confounding

On occasion, **all plausible confounding** from observational studies or randomised trials may be working to **reduce the demonstrated effect** or **increase the effect if no effect was observed**.
For example, if only sicker patients receive an experimental intervention or exposure, yet they still fare better, it is likely that the actual intervention or exposure effect is larger than the data suggest.

Only studies with no important threats to validity should be upgraded.

» Compare diagram.

Example 1

A rigorous systematic review of observational studies including a total of 38 million patients demonstrated higher death rates in private for-profit versus private not-for-profit hospitals (Devereaux 2004). One possible bias relates to different disease severity in patients in the two hospital types. It is likely, however, that patients in the not-for-profit hospitals were sicker than those in the for-profit hospitals. Thus, to the extent that residual confounding existed, it would bias results against the not-for-profit hospitals. The second likely bias was the possibility that higher numbers of patients with excellent private insurance coverage could lead to a hospital having more resources and a spill-over effect that would benefit those without such coverage. Since for-profit hospitals are likely to admit a larger proportion of such well-insured patients than not-for-profit hospitals, the bias is once again against the not-for-profit hospitals. Because the plausible biases would all diminish the demonstrated intervention effect, one might consider the evidence from these observational studies as moderate rather than low quality.

Example 2

A parallel situation exists when observational studies have failed to demonstrate an association but all plausible biases would have increased an intervention effect. This situation will usually arise in the exploration of apparent harmful effects. For example, because the hypoglycaemic drug phenformin causes lactic acidosis, the related agent metformin is under suspicion for the same toxicity. Nevertheless, very large observational studies have failed to demonstrate an association (Salpeter S, Greyber E, Pasternak G, Salpeter E. Risk of fatal and nonfatal lactic acidosis with metformin use in type 2 diabetes mellitus. Cochrane Database of Systematic Reviews 2007, Issue 4. Art No: CD002967.). Given the likelihood that clinicians would be more alert to lactic acidosis in the presence of the agent and overreport its occurence, one might consider this moderate, or even high quality evidence refuting a causal relationship between typical therapeutic doses of metformin and lactic acidosis.

8.7.1.9 Dose-response gradient

About Dose-response Gradient

The presence of a **dose-response gradient** may increase our confidence in the findings of observational studies and thereby increase the quality of evidence.

Only studies with no threats to validity (not downgraded for any reason) can be upgraded.

» Compare diagram.

Example

The observation that, in patients receiving anticoagulation with warfarin, there is a dose response gradient between higher levels of the international normalized ratio (INR), an indicator of the degree of anticoagulation, and an increased risk of bleeding increases our confidence that supratherapeutic anticoagulation levels increase bleeding risk.

8.7.1.10 Examples of quality of evidence grades

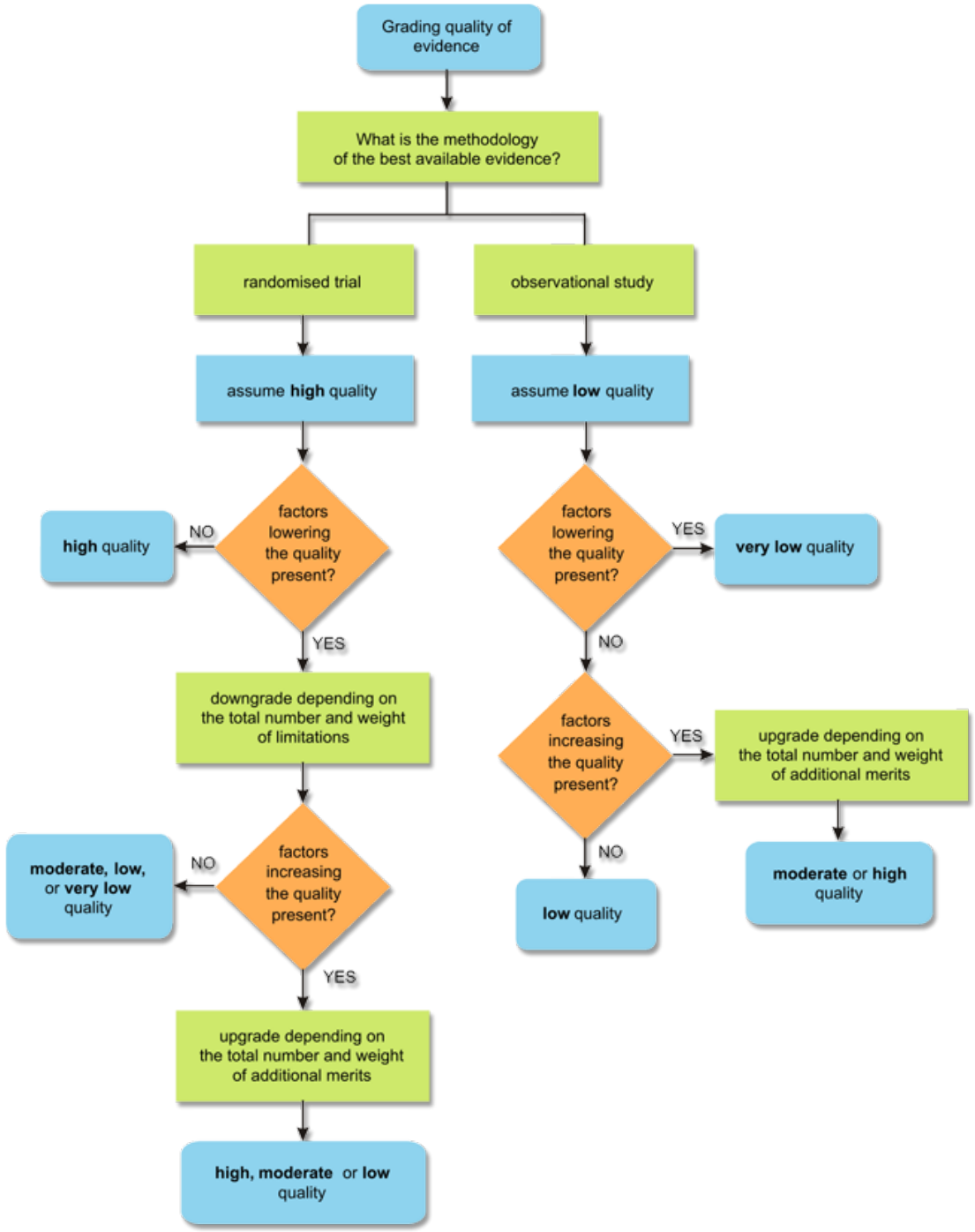
Examples of Grades of the Quality of Evidence

Example Study Type	Quality of Evidence
RCT(s) with no limitations, consistent, precise, and directly applicable results without evidence of reporting bias	**high**
RCT(s) with important limitations	**moderate** (downgraded from high)
RCT(s) with very serious limitations	**low** (downgraded two levels from high)
RCT(s) with very serious limitations and inconsistent results	**very low** (downgraded three levels from high)
Observational studies with no threats to validity yielding very large effects	**high** (upgraded two levels from low)
Observational studies with no threats to validity and evidence of a dose-response gradient	**moderate** (upgraded from low)
Observational studies with no threats to validity	**low**
Observational studies with uncertainty about the directness of results	**very low**
Unsystematic observations (case series or case reports)	**very low**

8.7.2 Quality of evidence diagram

Assessing the Quality of Evidence

When assessing quality of evidence you may follow this diagram.



You may also download and print the above diagram.
[You will need Acrobat Reader in order to view this document]

8.7.3 Overall quality of evidence

About the Overall Quality of Evidence

The overall quality of evidence is a combined grade of the quality of evidence across many outcomes considered critical for a recommendation.

FOR GUIDELINE PANELS

Guideline panels **have to determine the overall quality of evidence** across all the critical outcomes essential to a recommendation they make. Guideline panels usually provide a single grade of quality of evidence for every recommendation, but the strength of a recommendation usually depends on evidence regarding not just one, but a number of patient-important outcomes and on the quality of evidence for each of these outcomes.

When determining the overall quality of evidence across outcomes:

1. Consider **only** those outcomes that are deemed **critical**.
2. If the quality of evidence differs across critical outcomes and
   - outcomes point in different directions — **towards benefit and towards harm** — the **lowest quality of evidence** for any of the critical outcomes determines the overall quality of evidence
   - all outcomes point in the same direction — **towards either benefit or harm** — the **highest quality of evidence** for a critical outcome that by itself would suffice to recommend an intervention determines the overall quality of evidence. However, if the balance of the benefits and downsides is uncertain, then the grade of the critical outcome with the lowest quality grading should be assigned.

FOR AUTHORS OF SYSTEMATIC REVIEWS

Authors of systematic reviews **do not grade the overall quality of evidence** across outcomes.
Because systematic reviews do not – or at least should not – make recommendations, authors of systematic reviews rate the quality of evidence only for each outcome separately.

EXAMPLE

Consider administration of selective digestive decontamination (SDD) in intensive care unit patients. Several systematic reviews of high-quality randomised trials suggest a decrease in the incidence of infections and, likely, the mortality of ventilated patients receiving SDD. The quality of evidence on the effect of SDD on the emergence of bacterial antibiotic resistance and its clinical relevance is much less clear. One might reasonably grade the evidence about this feared potential adverse effect as low quality. If those making a recommendation felt that these downsides of therapy were critical, the overall grade of the quality of evidence for SDD would be low. If guideline panel felt that the emergence of bacterial antibiotic resistance was important but not critical, the grade for an overall quality of evidence would be high.

Copyright © 2004–2009 GRADE Working Group

8.8 Going from evidence to recommendations

Going from Evidence to Recommendations

- Definition of the strength of a recommendation
- Factors determining the strength of a recommendation
- Wording of a recommendation
- Symbolic representation of the strength of a recommendation
- Interpretation of a recommendation

8.8.1 Strength of a recommendation

About Strength of a Recommendation

The **strength of a recommendation** reflects the extent to which a guideline panel is **confident that desirable effects of an intervention outweigh undesirable effects**, or vice versa, across the range of patients for whom the recommendation is intended.

GRADE specifies only **two categories** of the strength of a recommendation. While GRADE suggests using the terms strong and **weak** recommendations, those making recommendations may choose different wording to characterize the two categories of strength.

For a guideline panel or others making recommendations to offer a strong recommendation they have to be **certain** about the various factors that influence the strength of a recommendation. The panel also should have the relevant information at hand that supports a clear balance towards either the desirable effects of an intervention (to recommend an action) or undesirable effects (to recommend against an action).
When a guideline panel is **uncertain** whether the balance is clear or when the relevant information about the various factors that influence the strength of a recommendation is not available, a guideline panel should be more cautious and in most instances it would opt to make a weak recommendation.

To aid interpretation GRADE suggests implications of strong or weak recommendations that follow from the recommendations.

8.8.1.1 Strong recommendation

About Strong Recommendations

Strong recommendation

A strong recommendation is one for which guideline panel is confident that the desirable effects of an intervention outweigh its undesirable effects (strong recommendation for an intervention) or that the undesirable effects of an intervention outweigh its desirable effects (strong recommendation against an intervention).

A strong recommendation implies, that most or all individuals will be best served by the recommended course of action.

Example 1

Early anticoagulation in patients with deep venous thrombosis for the prevention of pulmonary embolism; antibiotics for the treatment of community acquired pneumonia; quit smoking to prevent adverse consequences of tobacco smoke exposure; bronchodilators in patients with known COPD.

Example 2

Thromboprophylaxis reduces the incidence of venous thromboembolism in immobile, hospitalized severely ill medical patients. Careful thromboprophylaxis has minimal side effects and relatively low cost while being very effective at preventing deep venous thrombosis and its sequelae. Peoples' values and preferences are such that virtually all patients admitted to a hospital would, if they understood the choice they were making, opt to receive some form of thromboprophylaxis. Those making recommendations can thus offer a strong recommendation for thromboprophylaxis for patients in this setting.

8.8.1.2 Weak recommendation

About Weak Recommendations

Weak recommendation

A weak recommendation is one for which the desirable effects probably outweigh the undesirable effects (weak recommendation for an intervention) or undesirable effects probably outweigh the desirable effects (weak recommendation against an intervention) but appreciable uncertainty exists.

A weak recommendation implies, that not all individuals will be best served by the recommended course of action. There is a need to consider more carefully than usual individual patient's circumstances, preferences, and values.

Example 1

LVRS in patients with severe (upper lobe-predominant) emphysema and low exercise capacity; indefinite anticoagulation in patients with idiopathic VTE.

Example 2

Consider a 40 year-old man who has suffered an idiopathic deep venous thrombosis (DVT) followed by treatment with adjusted dose warfarin for one year to prevent recurrent DVT and pulmonary embolism. Continuing on standard-intensity warfarin beyond the treatment of one year will reduce his absolute risk for recurrent DVT by more than 7% per year for several years. The burdens of treatment include taking a warfarin pill daily, keeping dietary intake of vitamin K constant, monitoring the intensity of anticoagulation with blood tests, and living with the increased risk of both minor and major bleeding. Patients who are very averse to a recurrent DVT would consider the benefits of avoiding DVT worth the downsides of taking warfarin. Other patients are likely to consider the benefit not worth the harms and burden.

Example 3

Lung volume reduction surgery (LVRS) for severe emphysema offers another example of an intervention in which patient preferences and values play a central role in making treatment recommendations and decisions. Results of the only large-scale RCT to date indicate that lung resection, when combined with medical therapy, does not affect overall survival, although exercise capacity, quality of life, and other functional outcomes at 2 years are improved compared to medical therapy alone (12). However, surgery increases the risk of short-term mortality (5.2 vs. 1.5% at 90-days). In addition, the salutatory effects of surgery on functional outcomes appear to diminish with time. Thus, whereas some patients would be enthusiastic about undergoing LVRS because of the anticipated benefit in exercise capacity and quality of life, others who fear the risk of higher mortality in the early post-surgical phase may be less so. As in the example of anticoagulation for DVT, fully informed patients who are offered LVRS for severe emphysema are likely to make different choices regarding this procedure; guideline panels should therefore offer this treatment as a weak recommendation. Recommendations for or against LVRS may further differ by subgroups because secondary

analyses suggest that outcomes are highly variable across subgroups defined by the anatomical distribution of emphysema and maximal exercise capacity prior to surgery.

Example 4

A systematic review of randomized trials suggests that in 1,000 patients with ST elevation myocardial infarction who are receiving thrombolytic therapy and aspirin and who are treated with heparin (versus no treatment with heparin) 5 fewer will die, 3 fewer will have reinfarction, and 1 fewer will have a pulmonary embolus, while 3 more will have major bleeds. Further, these estimates are not precise, and the advantage in decreased infarctions may be lost after six months. The small, imprecise and possibly transient benefit leaves us less confident about any recommendation to use heparin in this situation. Hence, the recommendation is likely to be weak.

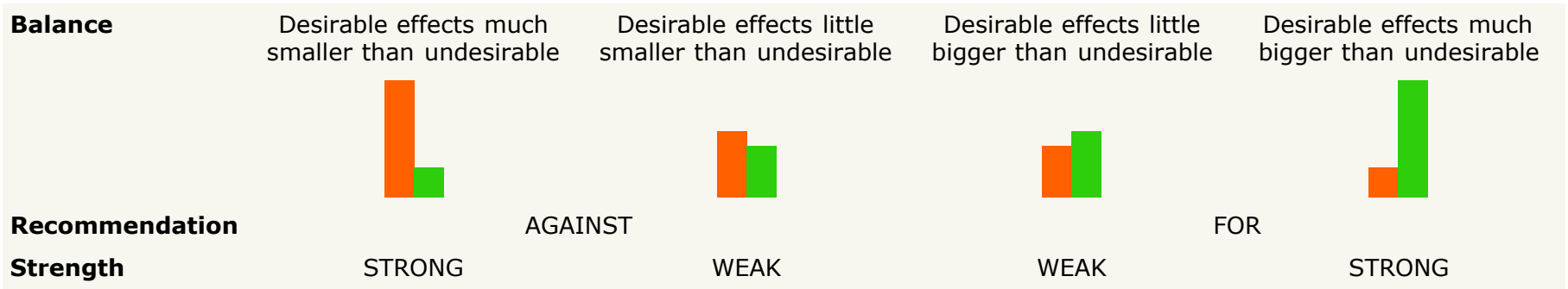
8.8.2 Balance of desirable and undesirable consequences

About the Balance of Desirable and Undesirable Consequences

The strength of a recommendation depends on the **balance** between **desirable consequences** (benefits) and **undesirable consequences** (downsides — harms, burden, and cost) of an intervention.

If **benefits outweigh downsides**, guideline panels will recommend that clinicians offer a diagnostic or therapeutic strategy to appropriately chosen patients.
If **downsides outweigh benefits**, guideline panels will recommend against the implementation of such a strategy.

Recommendations for or against interventions may be: strong or weak.



Guideline panels will usually recommend the management option that results in greater net benefit.

Desirable effects can include beneficial health outcomes (e.g. improved health related quality of life, fewer coronary events, or fewer hospitalizations), less burden and savings. **Undesirable effects** can include harms, more burden, and costs. Burdens are the demands of adhering to a recommendation that patients or caregivers (e.g. family) may dislike, such as having to take medication or the inconvenience of going to the doctor's office.

Those making recommendations should consider a number of factors when balancing the desirable and the undesirable effects. We will describe these factors in the next sections on:

- importance of outcomes
- baseline risk of outcomes
- relative and absolute effect of an intervention
- precision of the estimates of the effects
- cost

Related

» How guideline panels can put this information together

8.8.2.1 Importance of outcomes

About the Importance of Outcomes and Strength of Recommendations

Guideline panels should, in general, make **stronger recommendations** for interventions that increase the probability of beneficial outcomes or decrease the risk of adverse outcomes with **high** patient importance.

The importance of outcomes should be included in the considerations before making recommendations. Guideline panel should seek evidence about the actual and relative values that patients place on outcomes. They should also look for evidence about variability in preferences and values in patients and other stakeholders. If values and preferences vary widely a strong recommendation becomes less likely.

Example 1

Short-term aspirin reduces the relative risk of death after myocardial infarction by approximately 25%. Aspirin has minimal side effects and very low cost. Peoples' values and preferences are such that virtually all patients suffering a myocardial infarction would, if they understood the choice they were making, opt to receive aspirin. Guideline panels can thus offer a strong recommendation for aspirin administration in this setting.

Example 2

One needs to expose four patients to a respiratory rehabilitation program for one patient to gain a small but important improvement in dyspnea in daily life. In low-risk patients who have experienced a myocardial infarction, one might need to treat 100 patients with agents such as aspirin, ß-blockers, angiotensin-converting enzyme inhibitors, or statins, to extend the life of one patient. Despite the much higher number needed to treat, since we value prolongation of life more highly than relieving dyspnea, the latter intervention may warrant a stronger recommendation.

Example 3

One can compare (a) 5 patients with gastroesophageal reflux disease and chronic cough might need to be treated with a proton pump inhibitor for 1 patient to an uncertain benefit of cough reduction with (b) 10 patients with acute respiratory distress syndrome (ARDS) who one might need to treat with a low tidal volume ventilation strategy to prevent a premature death. Despite the higher number needed to treat (NNT) in the ARDS patient, since patients would value prolongation of life more highly than relieving cough, all else being equal, the latter intervention could warrant a stronger recommendation.

Example 4

Consider a 40 year-old man who has suffered an idiopathic deep venous thrombosis (DVT) followed by treatment with adjusted dose warfarin for one year to prevent recurrent DVT and pulmonary embolism. Continuing on standard-intensity warfarin beyond the treatment of one year will reduce his absolute risk for recurrent DVT by more than 7% per year for several years. The burdens of treatment include taking a warfarin pill daily, keeping dietary intake of vitamin K constant, monitoring the intensity of anticoagulation with blood tests, and living with the increased risk of both minor and major bleeding. The initial choice made by the patient to accept adjusted dose warfarin for one year versus shorter periods (less than 3 months) for the prevention of DVT recurrence or other adverse outcomes in patients with initial DVT illustrates several of the factors that will influence the strength of a recommendation.

Copyright © 2004–2009 GRADE Working Group

8.8.2.2 Baseline risk

About the Baseline Risk

In general, **the higher the** baseline risk, the greater is the magnitude of benefit and **the more likely the recommendation will be strong**.

If the baseline risk of an outcome significantly influences the magnitude of benefit or harm from an intervention or exposure, and there are specific populations of patients at high or low risk of an outcome, then guideline panels should consider making separate recommendations for these different populations.

Example 1

Consider the choice of adjusted-dose warfarin vs aspirin for the prevention of stroke in patients with atrial fibrillation. A systematic review and metaanalysis8 found a relative risk reduction (RRR) of 46% in all strokes with warfarin vs aspirin. Consider a 65-year old patient with atrial fibrillation and no other risk factors for stroke. This individual's risk for stroke in the next year is approximately 2%. Therapy with dose-adjusted warfarin can, relative to aspirin, reduce the risk to approximately 1%. An older person with additional risk factor for stroke has a risk for the stroke over the next year of approximately 8%. Therapy with dose adjusted Warfarin can, relative o aspirin reduce the risk to approximately 4%. The use of Warfarin may warrant a stronger recommendation in a population similar to the second patient.

Example 2

Consider a 65 year-old patient with mild COPD and frequent exacerbations for whom inhaled corticosteroids are a treatment option. This individual's risk for suffering an exacerbation in the next year may be 20%. Considering the relative risk reduction of inhaled corticosteroids for reducing exacerbations (RR 0.76, 95% CI 0.72 to 0.80) and this baseline risk, one can derive a simplified absolute magnitude of the effect (19). Inhaled corticosteroids, relative to placebo, will reduce the absolute risk by approximately 4.8% (= 20% - (0.76 x 20%)). Some patients who are very averse to experiencing an exacerbation may consider the downsides of inhaled corticosteroids (thrush, fracture risk, burden of inhalers) well worth it. Given the relative narrow confidence interval that follows from the confidence interval around the relative risk reduction one could make a strong recommendation for using inhaled corticosteroids if all patients were equally adverse to exacerbations. More patients are, however, likely to consider the benefit not worth the harms and burden of taking inhalers if their baseline risk is lower. For instance, if the baseline risk for an exacerbation is 5%, the absolute risk reduction is only 1.2% (= 5% - (0.76 x 5%)) but the possible harms and burden remain unchanged. Fewer patients with lower baseline risk would make the choice of taking inhaled steroids. When, across the range of patient values, fully informed patients are liable to make different choices, guideline panels should offer weak recommendations and explain the rationale for their recommendation.

8.8.2.3 Magnitude of effect

About Relative and Absolute Effect and Strength of Recommendations

Large relative effects

- If large relative effects of an intervention consistently point in the **same direction** — towards benefits or towards harms and burdens — they are likely to lead to a **strong** recommendation.
- If large relative effects of an intervention point in **opposite directions** — large benefits accompanied by large risk of adverse effects — the recommendation is more likely to be **weak**.

Example

Oral anticoagulant versus placebo in patients with atrial fibrillation leads to a greater reduction in the risk of stroke (RRR: 68%) than administering clopidogrel versus aspirin in patients

with transient ischemic attack (RRR: 8.7%). A recommendation for using oral anticoagulant in patients with atrial fibrillation is therefore more likely to be strong than a recommendation to use clopidogrel in patients with transient ischemic attack.

Large absolute effects

Large absolute effects are more likely to lead to a strong recommendation, than small absolute effects.

Example

In patients with a similar risk of stroke of 4% per year, administering warfarin versus aspirin to patients with atrial fibrillation (risk difference: 2%) leads to a greater absolute stroke reduction per year than administering clopidogrel versus aspirin to patients with previous transient ischaemic attack (risk difference: 0.3%).

8.8.2.4 Precision of estimate

About Precision of the Estimates of the Effects and Strength of Recommendations

The more precise are the estimates of the effect of an intervention, the more likely the recommendation will be strong.

Example

The estimate of the effect of ASA versus placebo on the prevention of stroke in patients with atrial fibrillation has a wider confidence interval than the estimate of the effect of ASA versus placebo in patients with transient ischaemic attacks.

8.8.2.5 Cost (resource use)

About Cost (Resource Use)

Cost may be considered just another potentially important outcome – like mortality, morbidity, and quality of life – associated with alternative ways of managing patient problems. In addition to these clinical outcomes, however, an intervention may increase costs or decrease costs.

Special considerations when incorporating resources use (cost) in recommendations

- What are the differences between costs and other outcomes?
- Which perspective to take?
- Which resource implications to include?

- How to make judgments about the quality of the evidence?
- How to present these implications?
- What is potential usefulness of a formal economic model?
- How to consider resource use in formulating recommendations?

8.8.2.5.1 Differences between costs and other outcomes

About Differences Between Costs and Other Outcomes

There are several differences between costs and other outcomes:

1. With costs the issue of who pays and who gains is most prominent.
2. Attitudes about the extent to which costs should influence the decision differ depending on who bears the cost.
3. Costs tend to vary widely across jurisdictions and over time.
4. People have different perspectives on the envelope in which they are considering opportunity costs.
5. Resource allocation is a far more political issue than consideration of other outcomes.

Despite these differences, approaches to cost (resource use) are similar to other outcomes:

- guideline panels need to consider only important resource implications
- decision makers require an estimate of the difference between treatment and control
- guieline panels must make explicit judgments about the quality of evidence regarding incremental resource use.

Differences between costs and other outcomes

1. With costs the issue of who pays and who gains is most prominent.
For most outcomes other than costs, it is clear that the patient and, secondarily, the patient's family gains the advantages, and has to live with the disadvantages (this is not true of all outcomes – with vaccinations the entire community benefits from the herd effect, or widespread use of antibiotics may have down-stream adverse consequences of drug resistance). Health care costs are often borne by the society as a whole. Even within a society, who bears the cost may differ depending on the patient's age or situation.

2. Attitudes about the extent to which costs should influence the decision differ depending on who bears the cost.
If costs are borne by the government, or a third party payer, some would argue that the physician's responsibility to the patient means that costs should not influence the decision. On the other hand, a clinicians' responsibility when caring for a patient is discharged in a broader context: resources that are used for an intervention cannot be used for something else and can affect the ability of the health system to best meet the needs of those it serves.

3. Costs tend to vary widely across jurisdictions or even within jurisdictions, and over time.
Costs of drugs are largely unrelated to the costs of production of those drugs, and more to marketing decisions and national policies. Hospitals or health maintenance organizations may, for instance, negotiate special arrangements with pharmaceutical companies for prices substantially lower than are available to patients or other providers. Even when resource use remains the same, the resource implications may vary widely across jurisdictions. Costs can also vary widely over time (e.g. when a drug comes off patent or a new, cheaper technology becomes available). The large variability in costs over time and jurisdictions requires that guideline panels formulate health care questions as specific as possible when bringing cost into the equation. The choice of comparator can be a particular problem in economic analyses. If the choice of the comparator is inappropriate (for instance, no treatment rather than an alternative though less effective intervention) conclusions may be misleading. Even when resource use remains the same, the resource implications may vary widely across jurisdictions. A year's supply of a very expensive drug may pay a nurse's salary in the United States, six nurses' salaries in Poland, and 30 nurses' salaries in China. Thus, what one can buy with the resources saved if one foregoes purchase of the drug (the "opportunity cost") – and the health benefits achieved with those expenditures - will differ to a large extent.

4. People have different perspectives on the envelope in which they are considering opportunity costs.
A hospital pharmacy with a fixed budget considering purchase of an expensive new drug will have a clear idea of what that purchase will mean in terms of other medications the pharmacy cannot afford. People often assume the envelope is public health spending – funding a new drug or program will constrain resources for other public health expenditures. However, one may not be sure that refraining from that purchase really means that equivalent resources will be available for the health care system. Further, one may ask if the public health care is spending the correct envelope.

5. Resource allocation is a far more political issue than consideration of other outcomes.
Whether the guideline panel does or does not explicitly consider resource allocation issues, those politics may bear on a guideline panel's function through conflict of interest.

8.8.2.5.2 Perspective

About the Perspective Taken When Considering Cost

GRADE suggests that a broad perspective is desirable.

A recommendation could be intended for a very narrow audience, such as a single hospital pharmacy, an individual hospital or a health maintenance organization. Alternatively it could be intended for a health region, a country or an international audience.
Regardless of how narrow or broad the intended audience, guideline groups that choose to incorporate resource implications must be explicit about the perspective they are taking.

Alternatively a guideline may choose to take a societal perspective, and include all important resource implications, regardless of who bears the costs.

In a publicly funded health system the patient perspective would consider only resource implications that directly affect individual patients (e.g. out of pocket costs) and would ignore most of the costs generated (e.g. costs borne by the government). In European health care systems in which, for the most part, governments bear the cost of health care, expenses borne directly by patients will be minimal. A pharmacy perspective would ignore down-stream cost savings resulting for adverse events (e.g. stroke or myocardial infarction) prevented by a drug. A hospital perspective would ignore out-patient costs either incurred, or prevented. In the private sector, where disenrollment and loss of insurance can shift the burden of costs from one system to another, estimates of resource use should include the down-stream costs of all treated patients, not just those who remain in a particular health plan.
An even broader perspective, that of society, would include indirect costs or savings (e.g. lost wages). These are difficult to estimate and controversial because they assume that lost productivity will not be replaced by an individual who otherwise would be unemployed or underemployed, and implicitly place lower value on individuals not working (e.g. the retired). Taking a health systems perspective has another advantage. A comprehensive display of the resource use associated with alternative management strategies allows an individual or group – a patient, a pharmacy, or a hospital – to examine the relative merits of the alternatives from their particular perspective.

Clinicians seeing patients who are uncovered by either public or private insurance may need to help these individuals to make decisions taking into account their out of pocket costs. This is particularly true when clinical advantages and disadvantages are closely balanced, and there are substantial out of pocket costs. In these circumstances, if a guideline panel has used the GRADE approach and made evidence profiles available to the guideline users, clinicians can review evidence summaries and ensure that the patients' decision to accept the recommended management strategy is consistent with their values and preferences – either though communicating the information directly to the patient, or by finding out what the patients' situation and values and preferences are.

8.8.2.5.3 Resource implications considered

About Considered Resource Implications

Evidence profiles should always present resource use, not just monetary values.

We suggest that guideline developers document best estimates of resource use, not best estimate of costs. Costs are a function of resources expended and the cost per unit of resource. Given the wide variability in costs per unit, reporting only total costs across broad categories of resource expenditure leaves users without the information required to judge whether estimates of unit costs apply to their setting.

Users of guidelines will be best informed if the guideline developers specify resources consumed by alternate management strategies, because they can:

- judge whether the resource use reflects practice patterns in their setting
- focus on the items of most relevance to them
- ascertain whether the unit costs apply in their setting.

Unless resource use is specified, users in settings other than that on which the analysts focus cannot estimate the associated incremental costs of the intervention.

8.8.2.5.4 Quality of the evidence about cost

About the Quality of Evidence About Cost

Judging quality of evidence for resource use is similar to that for the other outcomes.

As with evidence of rare but serious adverse effects, evidence of resource use may come from different sources than evidence of health benefits. This may be the case both because trials of interventions do not fully report resource use, because the trial situation may not fully reflect the circumstances (thus the resource use) that we would expect in clinical practice, because the relevant resource use may extend beyond the duration of trial, and because resource use may vary substantially across settings.

For resource use that is reported in the context of trials, criteria for quality assessment are identical to that of other outcomes. Just as for other outcomes of a trial, the quality of evidence may differ across different resources. For example, drug use may be relatively easy to estimate, whereas use of health professionals' time may be more difficult, and the estimate of drug use may therefore be of higher quality.

8.8.2.5.5 Presentation of resource use

About Presentation of Resource Use

A balance sheet (evidence profile) should inform judgments about whether the net benefits are worth the incremental costs.

Example

The updated version of this handbook will include an example of a balance sheet.

8.8.2.5.6 Economic model

About Usefulness of Formal Economic Models

Formal economic modeling may – or may not - be helpful.

Formal economic modeling results in cost per unit benefit achieved: cost per natural unit, such as cost per stroke prevented (cost-effectivenes analysis) cost per quality-adjusted life year gained (cost-utility analysis) cost and benefits valued in monetary values (cost-benefit analysis). These summaries can be helpful for informing judgments. Unfortunately, many published cost-effectiveness analyses have a high probability of being flawed or biased, and are setting-specific.

Should guideline panels consider developing their own formal economic model?

Creating an economic model may be advisable if:

- guideline groups have the necessary expertise and resources
- difference in resources consumed by the alternative management strategies is large and therefore there is substantial uncertainty about whether the net benefits of an intervention are worth the incremental costs
- quality of available evidence regarding resource consumption is high and it is likely that a full economic model would help inform a decision
- implementing an intervention requires large capital investments, such as building new facilities or purchasing new, expensive equipment.

Modeling – while necessary for taking into account complexities and uncertainties in calculating cost per unit benefit – reduces transparency. Any model is only as good as the data on which it is based. When estimates of benefits, harms, or resources used come from low quality evidence, results of any economic modeling will be highly speculative.

Although criteria to assess the credence to give to results from statistical models of cost-effectiveness or cost-utility are available, these models generally include a large number of assumptions and varying quality evidence for the estimates that are included in the model. For these reasons, GRADE working group recommends not including cost-effectiveness or cost-utility models in evidence profiles. These models may, however, inform judgments of a guideline panel, or those of governments, or third part payers considering whether to include an intervention among their programs' benefits.

8.8.2.5.7 Consideration of resource use in recommendations

About Considering of Resource Use in Recommendations

Guideline panel may choose to explicitly consider or not to consider resource use in recommendations.

A guideline panel may legitimately choose to leave considerations of resource use aside, and offer a recommendation solely on the basis of other advantages and disadvantages of the alternatives being considered. Resource allocation must then be considered at the level of the ultimate decision-maker – be it the patient and healthcare professional, an organization (e.g. hospital pharmacy or a health maintenance organization), a third party payer, or a government. Guideline panels should be explicit about the decision to consider or not to consider resource utilization.

If guideline panel considers resource use it should, prior to bringing cost into the equation, first decide on the quality of evidence regarding other outcomes, and weigh up the advantages and disadvantages. Decisions regarding the importance of resource use issues will flow from this first step. For example, resource implications may be irrelevant if evidence of net health benefits is lacking. If advantages of an intervention far outweigh disadvantages, resource use is less likely to be important. Resource use usually becomes important when advantages and disadvantages are closely balanced.

GRADE approach suggests that panels considering resource use should offer only a single recommendation taking resource use into account. Panels should refrain from issuing two recommendations – one not taking resource use into account and a second doing so. Although this would have the advantage of explicitness on which GRADE places a very high value, GRADE working group is concerned that those with interests in dissemination of an intervention would effectively use only the recommendation ignoring resource implications as a weapon in their battle for funds (public funds, in particular).

8.8.3 Confidence in values and preferences

About Confidence in Values and Preferences

Uncertainty concerning values and preferences or their **variability** among patients may lower the strength of a recommendation.

There will always be advantages and disadvantages of alternative management strategies. and patients will always have to make a trade-off between them. Therefore how a guideline panel values particular benefits, risks, and inconvenience can be decisive to any recommendation, and the strength of that recommendation.

Given the very limited study the issue of values and preferences towards different outcomes has received, one may argue that there is always large uncertainty about values and preferences. On the other hand, there is some systematic research of values and preferences, and guideline panel members' experience with patients provides additional insight.

Description of values and preferences considered when making a recommendation

While it is ideal for clinicians to elicit patient preferences and values directly from patients, and for guideline panels to obtain values and preference estimates from population based studies, such studies are often unavailable. When value or preference judgments are particularly important for the interpretation of recommendations, **authors should describe the key values they have attributed in making a recommendation**.

Example 1

A systematic review and meta-analysis describes a relative risk reduction (RRR) of approximately 80% in recurrent DVT for prophylaxis beyond 3 months up to one year. This large effect supports a strong recommendation for warfarin. Furthermore, the relatively narrow 95% confidence interval (approximately 74 to 88%) suggests that warfarin provides a RRR of at least 74%, and further supports a strong recommendation. At the same time, warfarin is associated with an inevitable burden of keeping dietary intake of vitamin K relatively constant, monitoring the intensity of anticoagulation with blood tests, and living with the increased risk of both minor and major bleeding. It is likely, however, that most patients would prefer avoiding another DVT and accept the risk of a bleeding episode. As a result, almost all patients with high risk of recurrent DVT would choose taking warfarin for 3 to 12 months, suggesting the appropriateness of a strong recommendation. Thereafter, there may be an appreciable number of patients who would reject life-long anticoagulation.

Example 2

Providing a recommendation for use of inhaled corticosteroids in mild COPD would require a statement about the higher value assigned to the fewer exacerbations, the possible, but

uncertain slower rate of FEV1 decline, and questionable mortality reduction than avoiding the harms from thrush, reduced bone mineral density, increased fracture risk, the burden of using inhalers and the cost associated with therapy.

8.8.4 Factors determining the strength of a recommendation

About Factors Determining the Strength of a Recommendation

Four key factors influence the strength of a recommendation

- Balance between desirable and undesirable effects
- Quality of evidence
- Values and preferences
- Cost (resource utilization)

Guideline panels can use simple decision tools to facilitate decision making. The following table provides a brief explanation of the key factors that those conducting guideline panel meetings can use to inform the panel about the key factors.

Factors that can strengthen a recommendation	Comment
Balance between desirable and undesirable effects (not considering cost)	The larger the difference between the desirable and undesirable consequences, the more likely a strong recommendation warranted. The smaller the net benefit and the lower certainty for that benefit, the more likely weak recommendation warranted.
Quality of the evidence	The higher the quality of evidence, the more likely is a strong recommendation.
Values and preferences	The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely weak recommendation warranted.
Costs (resource utilization)	The higher the costs of an intervention – that is, the more resources consumed – the less likely is a strong recommendation warranted

Example

The following table provides an example of a decision table that guideline developers can use to make judgements about each of these factors when making single recommendations. These tables can serve for record keeping similar to evidence profiles.

Decisions about the strength of a recommendation Example: treatment of H5N1 patients (bird or avian flu) with oseltamivir

Factors that determine the strength of a recommendation	Decision	Explanation
Balance between desirable and undesirable effects (not considering cost) **Is there uncertainty about the balance between desirable and undesirable effects (not considering cost)?**	☑ YES ☐ NO	The benefits are uncertain because several important or critical outcomes where not measured or the effects are closely balanced. However, the potential benefit is very large despite potentially small relative risk reductions.
Quality of evidence **Is the quality of evidence is low or very low?**	☑ YES ☐ NO	The quality of evidence is very low
Values and preferences	YES	All patients and care providers would accept treatment for H5N1 disease given the risk of

Is there a lot of variability in values and preferences, or is there uncertainty about the actual values and preferences?	☐ ☑ NO	mortality and the few downsides of the treatment
Costs (resource allocation) **Is there uncertainty whether the net benefits are worth the cost?**	☐ YES ☑ NO	For treatment of sporadic patients the price is not too high.

Frequent "yes" answers will increase the likelihood of a weak recommendation.

8.8.5 Wording of a recommendation

About Wording of a Recommendation

Wording of a recommendation should offer clinicians as many indicators as possible for **understanding and interpreting** the strength of recommendations.

For **strong recommendations**, the GRADE working group has suggested adopting terminology, such as "**we recommend**..." or "**clinicians should**...".

For **weak recommendations**, the GRADE working group has suggested less definitive wording, such as "**we suggest**..." or "**clinicians might**...".

Whatever terminology guideline panels use to communicate the dichotomous nature of a recommendation, it is essential that they inform their users what the terms imply.

Guideline panels should describe **patients or population** (characterized by the disease and other identifying factors) for whom the recommendation is intended and a recommended **intervention** as specifically and detailed as possible.

Wording strong and weak recommendations is particularly important when guidelines are developed by international organizations and/or are intended for patients and clinicians in different regions, cultures, traditions, and usage of language. It is also crucial to explicitly and precisely consider wording when translating recommendations into different languages.

8.8.6 Symbolic representation

About Symbolic Representation of the Quality of Evidence and Strength of Recommendation

A variety of presentations of quality of evidence and strength of recommendations may be appropriate. Most guideline panels have used **letters and numbers** to summarize their recommendations. Because of highly variable use of numbers and letters – some organizations have chosen letters for quality of evidence and numbers for strength of recommendations, and some the opposite – this presentation is potentially very confusing.

Symbolic representations of the quality of evidence and strength of recommendations are appealing in that they are not burdened with this historical confusion. On the other hand, clinicians seem to be very comfortable with numbers and letters, which are particularly suitable for verbal communication, so there may be good reasons why organizations have chosen to use them.

The GRADE working group has decided to offer preferred symbolic representations, but users of guidelines based on the GRADE approach will often see numbers and letters being used to express the quality of evidence and strength of a recommendation.

Suggested representations of quality of evidence and strength of recommendations

Quality of Evidence	Symbol	Letter
High	⊕⊕⊕⊕	A
Moderate	⊕⊕⊕◯	B
Low	⊕⊕◯◯	C
Very low	⊕◯◯◯	D

Strength of Recommendation	Symbol	Number
Strong for an intervention	↑↑	1
Weak for an intervention	↑?	2
Weak against an intervention	↓?	2
Strong against an intervention	↓↓	1

8.8.7 Interpretation and implications of a recommendation

Interpretation and Implications of a Recommendation

The advantage of GRADE two categories of strength of recommendations is that it provides clear direction to patients, clinicians, and policy-makers.

Implications of strong and weak recommendations for different users of guidelines

	Strong	Weak
for patients	Most individuals in this situation would want the recommended course of action and only a small proportion would not.	The majority of individuals in this situation would want the suggested course of action, but many would not.
for clinicians	Most individuals should receive the recommended course of action. Adherence to this recommendation according to the guideline could be used as a quality criterion or performance indicator. Formal decision aids are not likely to be needed to help individuals make	Recognize that different choices will be appropriate for different patients, and that you must help each patient arrive at a management decision consistent with her or his values and preferences. Decision aids may well be useful helping individuals making decisions consistent

	decisions consistent with their values and preferences.	with their values and preferences.
for policy makers	The recommendation can be adapted as policy in most situations including for the use as performance indicators.	Policy making will require substantial debates and involvement of many stakeholders. Performance indicators would have to focus on the fact that adequate deliberation about the management options has taken place.
Example	Question: Should intranasal for treatment glucocorticosteroids be used Recommendation: We recommend intranasal glucocorticosteroids for treatment of allergic rhinitis in adults and children [strong recommendation based on high quality evidence]	Question: Should antigen avoidance diet be used in pregnant or breastfeeding women to prevent development of allergy in children? Recommendation: For pregnant or breastfeeding women, we suggest no antigen avoidance diet. [weak recommendation based on very low quality evidence]

Individualization of clinical decision-making in weak recommendations remains a challenge. Although clinicians always should consider patients' preferences and values, when they face weak recommendations they may have a more detailed conversations with patients than for strong recommendations to ensure that the ultimate decision is consistent with the patient's preferences and values.

Important notice

Clinicians, patients, third-party payers, institutional review committees, other stakeholders, or the courts should **never view recommendations as dictates**. Even strong recommendations based on high-quality evidence will not apply to all circumstances and all patients.
Users of guidelines may reasonably conclude that following some strong recommendations based on the high quality evidence will be a mistake for some patients. No clinical practice guideline or recommendation can take into account all of the often compelling unique features of individual patients and clinical circumstances. Thus, nobody charged with evaluating clinician's actions, should attempt to apply recommendations in rote or blanket fashion.

8.8.8 Applying GRADE approach in guideline panels

Group and Guideline Panel Processes in the Application of GRADE

The following section lists useful tips for those applying the GRADE approach with guideline panels.

The research base to inform the choice of strategy to ensure appropriate group processes is limited, however in addition to logical arguments there is also some empirical evidence in support of using formal consensus development methods rather than relying only on informal processes (Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. Health Technol Assess 1998;2(3):i-iv, 1-88. Pagliari C, Grimshaw J. Impact of group structure and process on multidisciplinary evidence-based guideline development: an observational study. J Eval Clin Pract 2002;8(2):145-53. Pagliari C, Grimshaw J, Eccles M. The potential influence of small group processes on guideline development. J Eval Clin Pract 2001;7(2):165-73). Having one or two collaborating group leaders (chairs) that facilitate the group process is essential.

It is essential to secure that all participants have a chance to be heard and have the opportunity of contributing to the outcome of the process of developing recommendations using the GRADE approach. Neglecting that every (voting) group member has the opportunity to be heard is a common misunderstanding among groups that develop guidelines, and many have therefore adopted specific strategies to ensure appropriate group processes. There are several methods of ensuring that opportunities for contribution are given. These methods include closed processes such as soliciting information and (secret) voting and open processes including asking groups to provide feedback, solicit agreement and public voting in meetings. The latter processes are more likely to be influenced by group member characteristics including voicing strong opinions or timid behaviour. Disagreements will exist in most group decisions. It is important that strategies exist for dealing with disagreements.

The application of the GRADE approach requires consensus on judgments, in particular on recommendations, or at least a predetermined majority [see below] during several steps (there are some suggested solutions for situations when consensus cannot be reached):

1. The process begins with the requirement to accept using the GRADE approach by all parties or panel members.
2. Agreement must be reached on what the critical and important but not critical outcomes are for a given health care question. This can be done by asking panel members to suggest and generate outcomes and subsequent rating of the important of outcomes in an anonymous process (e.g. before a guideline meeting) and using the mean or median of the ratings for identifying the critical outcomes that will be evaluated in the process. Thus, a statistical approach could be used. This approach reflects the RAND approach in the following table (Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. Health Technol Assess 1998;2(3):i-iv, 1-88).

Characteristics of various consensus development methods (from Murphy et al. 1998)

Consensus development Method	Mailed Questionnaires	Private decisions elicited	Formal feed-back of group choices	Face-to-face contact	Structured interaction	Aggregation Method
Informal	Not applied	Not applied	Not applied	Yes	Not applied	Implicit
Delphi Method	Yes	Yes	Yes	Not applied	Yes	Explicit
Nominal Group Technique	Not applied	Yes	Yes	Yes	Yes	Explicit
Rand Version	Yes	Yes	Yes	Yes	Yes	Explicit
Other methods: Structured discussion	Not applied	Not applied	Not applied	Yes	Yes	Implicit

3. Agreement on the available evidence is required. All guideline panel members must be in agreement on both the underlying evidence and the judgments required when rating the quality of evidence during the development of evidence profiles. This can be done by allowing panel members sufficient time to review evidence profiles and contribute evidence that may have been missed when producing an evidence profile or summary. Consensus processes should be used in which arguments are heard and disagreements discussed until agreement is obtained. Judgments on evidence profiles are subjective and it is important that these judgments are presented in transparent fashion to possibly resolve disagreement. Sometimes voting may be necessary to achieve a decision or to force a decision.
4. Agreement must then be reached about the overall quality of evidence across all outcomes. This can be done easily if the rules are explained carefully (i.e., the overall quality of evidence results from the outcome with the lowest quality of evidence that the group judges as critical). However, it is important to explain these rules to the group beforehand.
5. Agreement must be reached on the balance of benefits, harms, burdens and cost. This can be done by putting forward suggestions (e.g. from the chair or other designated group members) or requesting suggestions from the group for the balance on a specific health care question. Using a formal checklist (link to decision table) can help making this process transparent and explicit. Situations exist in which groups can not make a decision or come to agreement on whether a recommendation should be made because evidence is sparse. These latter situations should be rare.
6. Having achieved agreement or deliberated on the balance of benefits, harms, burden and cost a panel should decided about the strength of a recommendation (i.e. weak or strong). Voting may be necessary regarding the strength of a recommendation (especially if formal decision modeling is not available – a likely situation) or whether a recommendation should be made at all.
7. Voting is one method to forcing agreement. Results of votes can be made public in guideline documents. An alternative method is to label that agreement has not been reached.

Steps 3 – 6 can be best achieved through discussion preceding a group meeting, so that group members are informed in advance.

Conflicts may arise within groups at any time and the leader of the group will have an important role in trying to manage these. Dealing with conflict is usually a difficult task, and guideline panels should consider establishing routines to support groups in managing these.

9. Glossary of terms and concepts

Glossary

This glossary is partially based on the glossary of Cochrane Collaboration terms and the Users' Guides to the Medical Literature with permission.

Absolute risk reduction (ARR): Synonym of the **risk difference** (RD). The difference in the risk between two groups. For example, if one group has a 15% risk of contracting a particular disease, and the other has a 10% risk of getting the disease, the risk difference is five percentage points.

Baseline risk: synonym of 'control event rate' or 'control group risk'.

Bias: A systematic error or deviation in results or inferences from the truth. In studies of the effects of health care, the main types of bias arise from systematic differences in the groups that are compared (**selection bias**), the care that is provided, exposure to other factors apart from the intervention of interest (**performance bias**), withdrawals or exclusions of people entered into a study (**attrition bias**) or how outcomes are assessed (**detection bias**). Systematic reviews of studies may also be particularly affected by **reporting bias**, where a biased subset of all the relevant data is available.

Burden: Burdens are the demands that patients or caregivers (e.g. family) may dislike, such as having to take medication or the inconvenience of going to the doctor's office.

Case series: A study reporting observations on a series of individuals, usually all receiving the same intervention, with no control group.

Case report: A study reporting observations on a single individual. Also called: anecdote, case history, or case study.

Case-control study: An observational study that compares people with a specific disease or outcome of interest (cases) to people from the same population without that disease or outcome (controls), and which seeks to find associations between the outcome and prior exposure to particular risk factors. This design is particularly useful where the outcome is rare and past exposure can be reliably measured. Case-control studies are usually retrospective, but not always.

Categorical data: Data that are classified into two or more non-overlapping categories. Race and type of drug (aspirin, paracetamol, etc.) are examples of categorical variables.

Clinical practice guideline (CPG): A systematically developed statement for practitioners and participants about appropriate health care for specific clinical circumstances.

Cohort study: An observational study in which a defined group of people (the cohort) is followed over time. The outcomes of people in subsets of this cohort are compared, to examine people who were exposed or not exposed (or exposed at different levels) to a particular intervention or other factor of interest. A **prospective** cohort study assembles participants and follows them into the future. A **retrospective** (or historical) cohort study identifies subjects from past records and follows them from the time of those records to the present.

Comparison: intervention against which new intervention is compared, control group.

Confidence interval (CI): A measure of the uncertainty around the main finding of a statistical analysis. Estimates of unknown quantities, such as the RR comparing an experimental intervention with a control, are usually presented as a point estimate and a 95% confidence interval. This means that if someone were to keep repeating a study in other samples from the same population, 95% of the confidence intervals from those studies would contain the true value of the unknown quantity. Alternatives to 95%, such as 90% and 99% confidence intervals, are sometimes used. Wider intervals indicate lower precision; narrow intervals, greater precision.

Confounder: A factor that is associated with both an intervention (or exposure) and the outcome of interest. For example, if people in the experimental group of a controlled trial are younger than those in the control group, it will be difficult to decide whether a lower risk of death in one group is due to the intervention or the difference in ages. Age is then said to be a confounder, or a confounding variable. Randomisation is used to minimise imbalances in confounding variables between experimental and control groups. Confounding is a major concern in non-randomised studies.

Consumer (healthcare consumer): Someone who uses, is affected by, or who is entitled to use a health related service.

Context: The conditions and circumstances that are relevant to the application of an intervention, for example the setting (in hospital, at home, in the air); the time (working day, holiday, night-time); type of practice (primary, secondary, tertiary care; private practice, insurance practice, charity); whether routine or emergency. Also called **clinical situation**.

Continuous data: Data with a potentially infinite number of possible values within a given range. Height, weight and blood pressure are examples of continuous variables.

Control: In a controlled trial a control is a participant in the arm that acts as a comparator for one or more experimental interventions. Controls may receive placebo, no treatment, standard treatment, or an active intervention, such as a standard drug. In an observational study a control is a person in the group without the disease or outcome of interest.

Control Event Rate (CER): observed risk of the event in the control group. Synonyms: control group risk, baseline risk. The control group risk for an outcome is calculated by dividing the number of people with an outcome in control group by the total number of participants in the control group.

Critical appraisal: The process of assessing and interpreting evidence by systematically considering its validity, results, and relevance.

Desirable effect: A desirable effect of adherence to a recommendation can include beneficial health outcomes, less burden and savings.

Dose response gradient: The relationship between the quantity of treatment given and its effect on outcome.

Effect size (ES): A generic term for the estimate of effect of treatment for a study. Sometimes the term is used to refer to the standardized mean difference.

To facilitate understanding we suggest interpretation of the effect size offered by Cohen (Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed; 1988). According to this interpretation, an effect size or SMD of around:

- **0.2** is considered a **small** effect
- **0.5** is considered a **moderate** effect
- **0.8** or higher is considered a **large** effect.

Effectiveness: The extent to which an intervention produces a beneficial result under ideal conditions. Clinical trials that assess effectiveness are sometimes called pragmatic or management trials.

Efficacy: The extent to which an intervention produces a beneficial result under ideal conditions. Clinical trials that assess efficacy are sometimes called explanatory trials.

Estimate of effect: The observed relationship between an intervention and an outcome expressed as, for example, a number needed to treat, odds ratio, risk difference, risk ratio, relative risk reduction, standardised mean difference, or weighted mean difference.

External validity: The extent to which results provide a correct basis for generalisations to other circumstances. For instance, a meta-analysis of trials of elderly patients may not be generalisable to children. Also called **generalisability** or **applicability**.

Follow-up: The observation over a period of time of study/trial participants to measure outcomes under investigation.

Hazard ratio (HR): A measure of effect produced by a survival analysis and representing the increased risk with which one group is likely to experience the outcome of interest. For example, if the hazard ratio for death for a treatment is 0.5, then we can say that treated patients are likely to die at half the rate of untreated patients.

Intention to treat analysis (ITT): A strategy for analysing data from a randomised controlled trial. All participants are included in the arm to which they were allocated, whether or not they received (or completed) the intervention given to that arm. Intention-to-treat analysis prevents bias caused by the loss of participants, which may disrupt the baseline equivalence established by randomisation and which may reflect non-adherence to the protocol. The term is often misused in trial publications when some participants were excluded.

Internal validity: The extent to which the design and conduct of a study are likely to have prevented bias. Variation in methodological quality can explain variation in the results of studies. More rigorously designed (better quality) trials are more likely to yield results that are closer to the truth.

Intervention: The process of intervening on people, groups, entities, or objects in an experimental study. In controlled trials, the word is sometimes used to describe the regimens in all comparison groups, including placebo and no-treatment arms.

Mean difference (MD): the 'difference in means' is a standard statistic that measures the absolute difference between the mean value in the two groups in a clinical trial. It estimates the amount by which the treatment changes the outcome on average. It can be used as a summary statistic in meta-analysis when outcome measurements in all trials are made on the same scale. Previously referred to as weighted mean difference (WMD).

Meta-analysis: The statistical combination of results from two or more separate studies.

Minimally important difference (MID): The smallest difference in score in the outcome of interest that informed patients or informed proxies perceive as important, either beneficial or harmful, and that would lead the patient or clinician to consider a change in the management.

Number needed to treat (NNT): An estimate of how many people need to receive a treatment before one person would experience a beneficial outcome. For example, if you need to give a stroke prevention drug to 20 people before one stroke is prevented, then the number needed to treat to benefit for that stroke prevention drug is 20. It is estimated as the reciprocal of the risk difference.

Number needed to harm (NNH): A number needed to treat to benefit associated with a harmful effect. It is an estimate of how many people need to receive a treatment before one more person would experience a harmful outcome or one fewer person would experience a beneficial outcome.

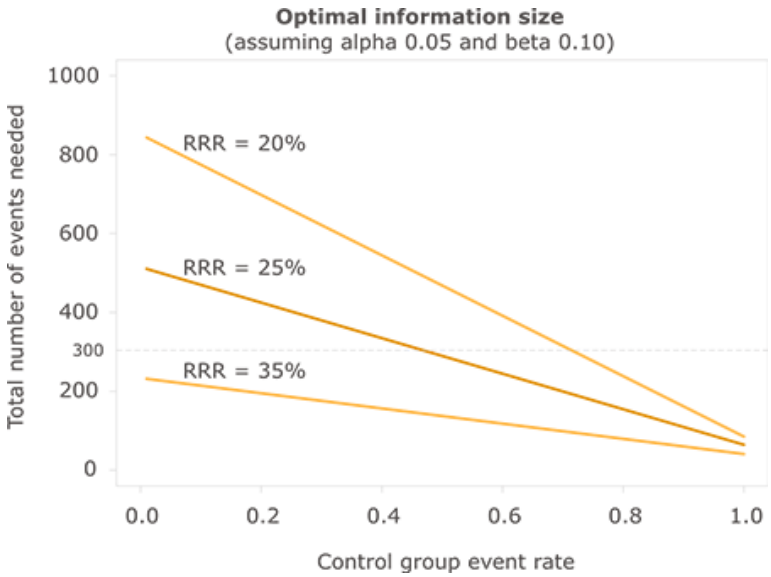
Observational study: A study in which the investigators do not seek to intervene, and simply observe the course of events. Changes or differences in one characteristic (e.g. whether or not people received the intervention of interest) are studied in relation to changes or differences in other characteristic(s) (e.g. whether or not they died), without action by the investigator. There is a greater risk of selection bias than in experimental studies.

Odds ratio (OR): The ratio of the odds of an event in one group to the odds of an event in another group. In studies of treatment effect, the odds in the treatment group are usually divided by the odds in the control group. An odds ratio of one indicates no difference between comparison groups. For undesirable outcomes an OR that is less than one indicates that the intervention was effective in reducing the risk of that outcome. When the risk is small, the value of odds ratio is similar to risk ratio.

Optimal information size (OIS): number of patients generated by a conventional sample size calculation for a single trial (Pogue and Yusuf, Controlled Clinical Trials, 1997;18:580-593).

Calculating the OIS for **dichotomous outcome** requires specifying:

- probability of detecting a false effect – type I error (a; usually 0.05)
- probability of detecting a true effect – power (usually 80% [power = 1 – type II error; ß; usually 0.20])
- realistic relative risk reduction (RRR; we suggest a default of 25%)
- control event rate (we suggest the median of the available trials, or the rate from a dominating trial, if it exists).



Calculating the OIS for **continuous variables** requires specifying:

- probability of detecting a false effect – type I error (a; usually 0.05)
- probability of detecting a true effect – power (usually 80% [power = 1 – type II error; ß; usually 0.20])
- realistic difference in means (Δ)
- appropriate standard deviation (SD) from one of the relevant studies (we suggest the median of the available trials or the rate from a dominating trial, if it exists).

Outcome: A component of a participant's clinical and functional status after an intervention has been applied, that is used to assess the effectiveness of an intervention.

Point estimate: The results (e.g. mean, weighted mean difference, odds ratio, risk ratio or risk difference) obtained in a sample (a study or a meta-analysis) which are used as the best estimate of what is true for the relevant population from which the sample is taken.

Population: The group of people being studied, usually by taking samples from that population. Populations may be defined by any characteristics e.g. geography, age group, certain diseases.

Precision: A measure of the likelihood of random errors in the results of a study, meta-analysis or measurement. The less random error the greater the precision. Confidence intervals around the estimate of effect from each study are one way of expressing precision, with a narrower confidence interval meaning more precision.

Quality of evidence: The extent to which one can be confident that an estimate of effect is correct.

Randomised controlled trial (RCT): An experimental study in which two or more interventions are compared by being randomly allocated to participants. In most trials one intervention is assigned to each individual but sometimes assignment is to defined groups of individuals (for example, in a household) or interventions are assigned within individuals (for example, in different orders or to different parts of the body).

Relative risk (RR): Synonym of risk ratio. The ratio of risks in two groups. In intervention studies, it is the ratio of the risk in the intervention group to the risk in the control group. A risk ratio of one indicates no difference between comparison groups. For undesirable outcomes, a risk ratio that is less than one indicates that the intervention was effective in reducing the

risk of that outcome.

Relative risk reduction (RRR): The proportional reduction in risk in one treatment group compared to another. It is one minus the risk ratio. If the risk ratio is 0.25, then the relative risk reduction is 1-0.25=0.75, or 75%.

Review Manager (RevMan): Software used for preparing and maintaining Cochrane systematic reviews. RevMan allows you to write ad manage systematic review protocols, as well as complete reviews, including text, tables, and study data. It can perform meta-analysis of the data entered, and present the results graphically. More information about RevMan.

Risk: The proportion of participants experiencing the event of interest. Thus, if out of 100 participants the event (e.g. a stroke) is observed in 32, the risk is 0.32. The control group risk is the risk amongst the control group. The risk is sometimes referred to as the event rate, and the control group risk as the control event rate. However, these latter terms confuse risk with rate.

Standardised mean difference (SMD): The difference between two estimated means divided by an estimate of the standard deviation. It is used to combine results from studies using different ways of measuring the same continuous variable, e.g. pain. By expressing the effects as a standardised value, the results can be combined since they have no units. Standardised mean differences are sometimes referred to as a d index.

Statistically significant: A result that is unlikely to have happened by chance. The usual threshold for this judgement is that the results, or more extreme results, would occur by chance with a probability of less than 0.05 if the null hypothesis was true. Statistical tests produce a p-value used to assess this.

Strength of a recommendation: the degree of confidence that the desirable effects of adherence to a recommendation outweigh the undesirable effects.

Surrogate outcome: Outcome measure that is not of direct practical importance but is believed to reflect an outcome that is important; for example, blood pressure is not directly important to patients but it is often used as an outcome in clinical trials because it is a risk factor for stroke and heart attacks. Surrogate outcomes are often physiological or biochemical markers that can be relatively quickly and easily measured, and that are taken as being predictive of important clinical outcomes. They are often used when observation of clinical outcomes requires long follow-up. Also called: intermediary outcomes or surrogate endpoints.

Systematic review: A review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review. Statistical methods (meta-analysis) may or may not be used to analyse and summarise the results of the included studies.

Undesirable effect: An undesirable effect of adherence to a recommendation can include harms, more burden, and costs.

10. References to articles about GRADE

References

1. Guyatt G.H., Oxman A.D., Vist G.E., Kunz R., Falck-Ytter Y., Alonso-Coello P., Schünemann H.J. **GRADE: an emerging consensus on rating quality of evidence and strength of recommendations.** BMJ, 2008;336:924-926. PMID: 18436948 [PubMed]
2. Guyatt G.H., Oxman A.D., Kunz R., Vist G.E., Falck-Ytter Y., Schünemann H.J. **What is "quality of evidence" and why is it important to clinicians?** BMJ, 2008;336:995-998. PMID: 18456631 [PubMed]
3. Guyatt G.H., Oxman A.D., Kunz R., Falck-Ytter Y., Vist G.E., Liberati A., Schünemann H.J. **Going from evidence to recommendations.** BMJ, 2008;336:1049-1051. PMID: 18467413 [PubMed]
4. Schünemann H.J., Oxman A.D., Brozek J., Glasziou P., Jaeschke R., Vist G.E., Williams J.W., Jr., Kunz R., Craig J., Montori V.M., Bossuyt P., Guyatt G.H. **Grading quality of evidence and strength of recommendations for diagnostic tests and strategies.** BMJ, 2008;336:1106-1110. PMID: 18483053 [PubMed]
5. Guyatt G.H., Oxman A.D., Kunz R., Jaeschke R., Helfand M., Liberati A., Vist G.E., Schünemann H.J. **Incorporating considerations of resources use into grading recommendations.** BMJ, 2008;336:1170-1173. PMID: 18497416 [PubMed]
6. Jaeschke R., Guyatt G.H., Dellinger P., Schünemann H.J., Levy M.M., Kunz R., Norris S., Bion J. **Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive.** BMJ, 2008;337:a744. PMID: 18669566 [PubMed]

7. Schünemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A, Fahy BF, Gould MK, Horan KL, Krishnan JA, Manthous CA, Maurer JR, McNicholas WT, Oxman AD, Rubenfeld G, Turino GM, Guyatt G; ATS Documents Development and Implementation Committee. **An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations.** Am J Respir Crit Care Med. 2006 Sep 1;174(5):605-614. PMID: 16931644 [PubMed]
8. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Schünemann HJ, Edejer TT, Varonen H, Vist GE, Williams JW Jr, Zaza S; GRADE Working Group. **Grading quality of evidence and strength of recommendations.** BMJ. 2004 Jun 19;328(7454):1490. PMID: 15205295 [PubMed]
9. Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, Hill S, Jaeschke R, Liberati A, Magrini N, Mason J, O'Connell D, Oxman AD, Phillips B, Schünemann H, Edejer TT, Vist GE, Williams JW Jr; GRADE Working Group. **Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system.** BMC Health Serv Res. 2005 Mar 23;5(1):25. PMID: 15788089 [PubMed]
10. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, Liberati A, O'Connell D, Oxman AD, Phillips B, Schünemann H, Edejer TT, Vist GE, Williams JW Jr; GRADE Working Group. **Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches.** BMC Health Serv Res. 2004 Dec 22;4(1):38. PMID: 15615589 [PubMed]
11. Schünemann H.J., Best D., Vist G., Oxman A.D. **Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations.** CMAJ, 2003;169:677-680. PMID: 14517128 [PubMed]
12. Schünemann HJ, Fretheim A, Oxman AD. **Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations.** BMC Health Research Policy and Systems, 2006;4:21. PMID: 17147810. [PubMed]
13. Guyatt G, Gutterman D, Baumann M, Adrizzio-Harris D, Hylek E, Phillips B, Raskob G, Schünemann H. Grading Strength of Recommendations and Quality of Evidence in Clinical Guidelines - Report from an ACCP working group. Chest, 2006;129:182-187. PMID: 16424429. [PubMed]
14. Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, Guyatt GH. **Chapter 12: Interpreting results and drawing conclusions.** In: Higgins JPT, Green S (editors), Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0 (updated February 2008). The Cochrane Collaboration, 2008. Available from www.cochrane-handbook.org.

11. Additional resources

Aditional Resources

Resources for authors of systematic reviews

The Cochrane Handbook

The Cochrane Handbook includes two principle chapters which provide information on how to create Summary of Findings tables using the information from Cochrane systematic reviews and GRADEing the evidence.
Part 2 Chapter 11: Presenting results and 'Summary of findings' tables
Part 2 Chapter 12: Interpreting results and drawing conclusions

General evidence-based medicine resources

The Cochrane Library

The Cochrane Library contains high-quality, independent evidence to inform healthcare decision-making. It includes reliable evidence from Cochrane and other systematic reviews, clinical trials, and more. Cochrane reviews bring you the combined results of the world's best medical research studies, and are recognised as the gold standard in evidence-based health care.

The Cochrane Handbook

The Cochrane Handbook for Systematic Reviews of Interventions (the Handbook) provides guidance to authors for the preparation of Cochrane Intervention reviews (including Cochrane

Overviews of reviews). The Handbook is updated regularly to reflect advances in systematic review methodology and in response to feedback from users.

Users' Guides to the Medical Literature

A complete set of Users' Guides to find, evaluate and use medical literature which were originally published as a series in the Journal of the American Medical Association (JAMA). Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice (Interactive) presents the sophisticated concepts of evidence-based medicine (EBM) in unique ways that can be used to determine diagnosis, decide optimal therapy, and predict prognosis. It also offers in-depth expansion of methodology, statistics, and cost issues that emerge in medical research.

Guideline specific resources

Improving the use of research evidence in guideline development (SERIES)

A series of 16 papers published in Health Research Policy and Systems in 2006, Volume 4, Issues 12 to 28 about guideline development. Topics are Guidelines for guidelines, Priority setting, Group composition and consultation process, Managing conflicts of interest, Group processes, Determining which outcomes are important, Deciding what evidence to include, Synthesis and presentation of evidence, Grading evidence and recommendations, Integrating values and consumer involvement, Incorporating considerations of cost-effectiveness, affordability and resource implications, Incorporating considerations of equity, Adaptation, applicability and transferability, Reporting guidelines, Disseminating and implementing guidelines, and Evaluation.

The AGREE instrument

The purpose of the Appraisal of Guidelines Research & Evaluation (AGREE) Instrument is to provide a framework for assessing the quality of clinical practice guidelines.

GRADE Working Group

The Grading of Recommendations Assessment, Development and Evaluation (short GRADE) Working Group began in the year 2000 as an informal collaboration of people with an interest in addressing the shortcomings of present grading systems in health care. Our aim is to develop a common, sensible approach to grading quality of evidence and strength of recommendation. References to articles about GRADE

Guidelines Advisory Committee

The Guidelines Advisory Committee (GAC) is an independent partnership of the Ontario Medical Association and the Ontario Ministry of Health and Long Term Care (MOHLTC). The GACs mission is to promote better health for the people of Ontario by encouraging physicians and other practitioners to use evidence-based clinical practice guidelines and clinical practices based on best available evidence. We identify, evaluate, endorse and summarize guidelines for use in Ontario.

National Guideline Clearing House

The National Guideline Clearinghouse™ (NGC) is a comprehensive database of evidence-based clinical practice guidelines and related documents. NGC is an initiative of the Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services.

National Library of Guidelines

The National Library of Guidelines is a collection of guidelines for the NHS. It is based on the guidelines produced by NICE and other national agencies. The main focus of the Library is on guidelines produced in the UK, but where no UK guideline is available, guidelines from other countries are included in the collection.

12. System requirements

System Requirements

GRADEprofiler and PC computers

GRADEprofiler runs on any computer with a Microsoft Windows 98®, XP, Vista or later operating system and Microsoft .NET Framework version 2.0 or 3.0.

Currently there is no plan to develop a GRADEprofiler version that would run under Linux.

GRADEprofiler and Apple computers

There are currently few options we can recommend for users with Apple Computers so they may use GRADEprofiler. These options are not comprehensive and the Mac users may choose other options if they are available.

- If you have a Mac with an **Intel** processor and **Microsoft® Windows istalled as an alternative system**, then you may run GRADEprofiler like on a PC.
- If you have a Mac with an **Intel** processor and **do not have Microsoft® Windows** installed as an alternative system, then you can purchase a third-party software such as VMware Fusion or Parallels Desktop for Mac that allows to install Windows and run Windows programs in Mac OS window. You would still have to obtain a copy of Microsoft® Windows and install it, but you will be able to use MacOS and Windows applications in parallel without rebooting to another system.
- If you have a Mac with **PowerPC** processor (reasonably new and fast one) and **Mac OS X**, then you may use the second solution described above.
- If you have a Mac with **PowerPC** processor and **Mac OS 9 or earlier**, then we do not know of any available solution at this time and most likely you will not be able to run GRADEprofiler on your computer.

13. Technical support

Technical support

GRADEprofiler is an evolving tool. Please check the Cochrane Collaboration GRADEpro website for updates and newer versions.

14. GRADE Working Group

About the GRADE Working Group

The **Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group** began in the year 2000 as an informal collaboration of more than 60 methodologists, clinicians, systematic reviewers, and guideline developers representing various organizations with the goal to address shortcomings of present grading systems in health care. The aim was to develop a common, sensible approach to grading quality of evidence and strength of recommendation. Based on shared experience, a critical review of other systems, and working through examples and applying the system in guidelines, the Working Group has developed the "GRADE approach" as a common, transparent and sensible method to grading quality of evidence and strength of recommendations. The GRADE working group has developed presentation formats evidence profiles.

Organizations that have endorsed or that are using GRADE*

Several organizations that are now using or endorsing GRADE approach in its original format or with minor modifications include:

	Guidelines for World Health Organization Guidelines	Example
	Endocrine Society Clinical Guidelines - USA	Example
	American College of Chest Physicians Guidelines - USA	Example
	UpToDate - Putting Clinical Information Into Practice - USA	
	Agenzia sanitaria regionale, Bologna - Italia	
	Ministry of Health and Long-Term Care, Ontario - Canada	Example
	Surviving Sepsis - International	
	Ärztliches Zentrum für Qualität in der Medizin - Germany	
	American Thoracic Society - USA	Example
	American College of Physicians - USA	Example
	The Cochrane Collaboration - International	

	Kidney Disease: Improving Global Outcome - International	Example
	European Society of Thoracic Surgeons - International	
	British Medical Journal - UK**	
JIDC	Journal of Infection in Developing Countries - International	
	Agency for Healthcare Research and Quality (AHRQ) - USA	Example
	Society of Critical Care Medicine (SCCM) - USA	
	National Institute for Clinical Excellence (NICE) - UK	
	Norwegian Knowledge Centre for the Health Services - Norway	
	The University of Pennsylvania Health System Center for Evidence-based Practice - USA	
	German Center for Evidence-based Nursing "sapere aude" - Germany	
	Evidence-based Nursing Südtirol, Alto Adige – Italy	
	Society for Vascular Surgery - USA	
	BMJ Clinical Evidence - UK	

	EBM Guidelines - Finland/International	
	Polish Institute for EBM - Poland	
	European Respiratory Society (ERS) - Europe	
	Japanese Society for Temporomandibular Joint - Japan	
	National Board of Health and Welfare - Sweden	
	COMPUS at The Canadian Agency for Drugs and Technologies in Health (CADTH) - Canada	Example
	Infectious Diseases Society of America - USA	

* some endorsements have included minor modifications, most commonly collapsing low and very low quality evidence into a single category
** groups submitting guidelines to the BMJ are encouraged to use GRADE

A complete and up-to-date list of participating organizations is posted on GRADE Working Group website.

15. Licence agreement

Licence agreement

By installing this application you are granted a licence to use this software provided that (1) you accept no liability of the producer and all its individual or institutional members or affiliates for any indirect, incidental, special or consequential damages (including without limitation any loss of data, interruption, computer failure or pecuniary loss) arising out of the use or inability to use this software, (2) you will not undertake the modification, creation of derivative works, translation, reverse engineering, decompiling, disassembling or hacking of this software or any

part thereof, and (3) you acknowledge and understand that the content of the information generated with this software is entirely the responsibility of the person who generated such a content.

16. GRADEpro credits and acknowledgements

Credits and Acknowledgements

The development of **GRADEprofiler**© (**GRADEpro**©) was supported by the European Commission Marie Curie Reintegration grant EU IGR42192 to Holger Schünemann, the Cochrane Collaboration, and the Norwegian Knowledge Centre for the Health Services.

Creation and Product Design:
Jan L. Brozek (brozek@mp.pl)
Andrew Oxman (oxman@online.no)
Holger J. Schünemann (schuneh@mcmaster.ca)

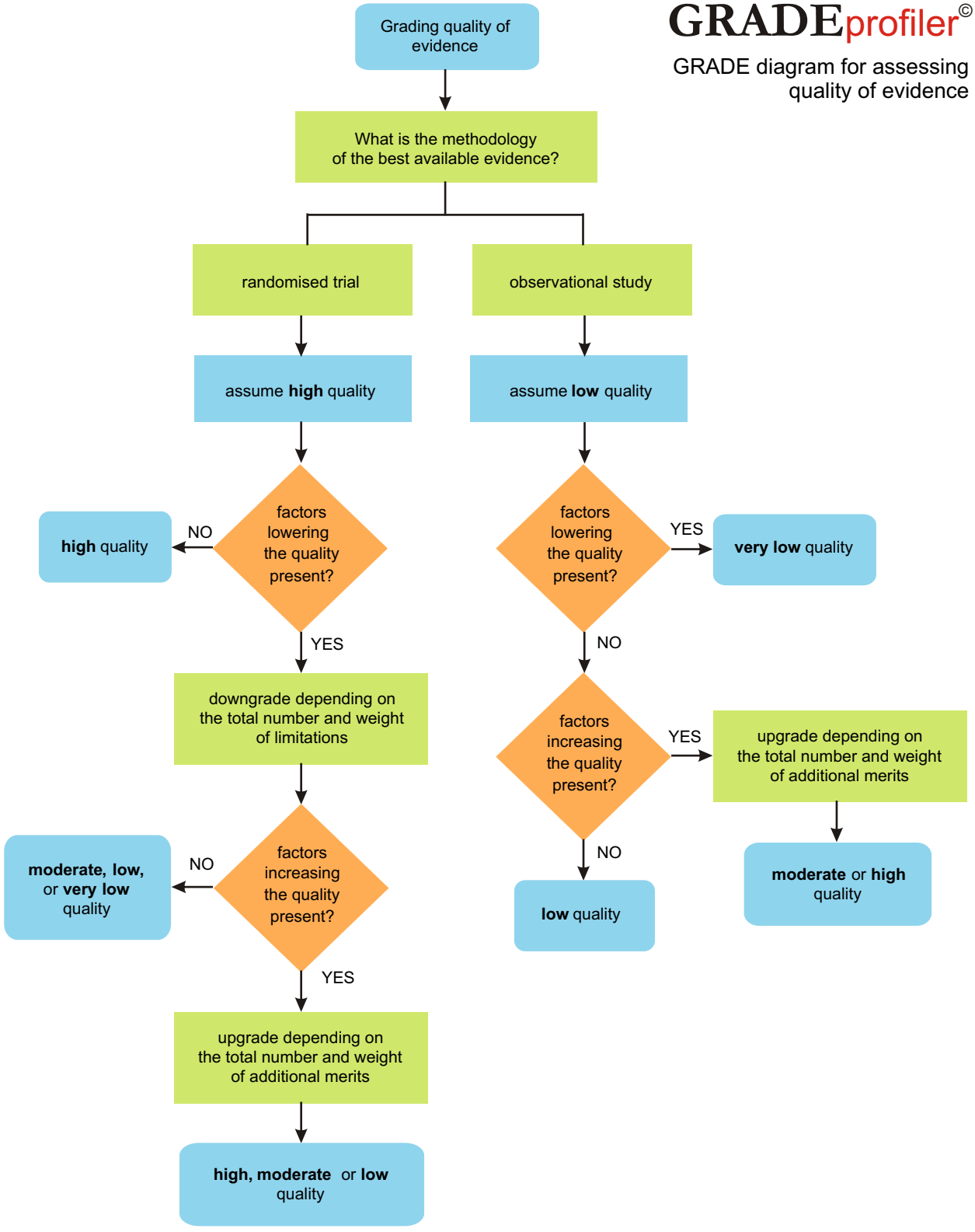
Critical reviewing:
Elie A. Akl (elieakl@buffalo.edu)
Nancy Santesso (santesso@uottawa.ca)

Programming:
Jakub Kapusta (jkapusta@mp.pl)

The GRADE Working Group (www.gradeworkinggroup.org) is an informal collaboration of people with the aim of developing and disseminating a sensible and transparent approach to grading quality of evidence and strength of recommendations in healthcare.

GRADEprofiler©

GRADE diagram for assessing quality of evidence



Factors that can reduce the quality of evidence

- Study limitations [⬇1 or 2 levels]
- Inconsistency of results [⬇1 or 2 levels]
- Indirectness of evidence [⬇1 or 2 levels]
- Imprecision [⬇1 or 2 levels]
- Reporting bias [⬇1 or 2 levels]

Factors that can increase the quality of evidence

- Large magnitude of effect [⬆1 or 2 levels]
- All plausible confounding would reduce observed effect [⬆1 level]
- Dose-response gradient [⬆1 level]