Handbook for Systematic Review Development

Version 2, February 2022

Author:
Elvira C. van Dalen (Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands) on behalf of the Systematic Review & Guideline Unit of the Princess Máxima Center for Pediatric Oncology

This handbook can be used by systematic review author teams for reviews developed in collaboration with the Systematic Review & Guideline Unit of the Princess Máxima Center for Pediatric Oncology (part of the Kremer Group). If you work in the Princess Máxima Center for Pediatric Oncology and want to start a systematic review please contact us at: E.C.vanDalen@prinsesmaximacentrum.nl

This Handbook should be cited as: Van Dalen EC, Handbook for Systematic Review Development, version 2, February 2022.

The author is grateful for the contributions of Renée Mulder, Jorrit van As and Leontien Kremer.

Contents

1 Introduction

1.1 Aim of the handbook

The main aim of this handbook is to provide guidance to systematic review author teams involved in preparing a systematic review in collaboration with the Systematic Review & Guideline Unit of the Princess Máxima Center for Pediatric Oncology (part of the Kremer Group), thereby improving the methodological quality of the systematic review. This can either be a Cochrane systematic review or another type of review outside Cochrane. There are different types of systematic reviews, with slight differences in methodology, but the key steps described in this handbook largely apply to all review types.

1.2 Systematic reviews

Systematic reviews provide a transparent and reproducible overview of the available evidence on a particular topic. This distinguishes systematic reviews from so-called narrative reviews, which often only express the opinion of experts and do not provide a complete picture of the available evidence. For example, in narrative reviews, a clear clinical question is lacking, no comprehensive search strategy has been performed and there is no quality assessment of the included results. In addition to insight into the available evidence for a particular question, systematic reviews provide insight into the possible lack thereof. This clarifies for which questions new studies are required before a well-founded recommendation can be made. Systematic reviews also form the basis for evidence-based guidelines. In this way, care for children with cancer and survivors can be optimized. However, to be able to do this it is essential that the reviews are of high quality.

Systematic reviews can be published in different journals. However, for Cochrane systematic reviews it is mandatory to publish in The Cochrane Library (for more information see point 1.3).

1.3 Cochrane

Founded in 1993, Cochrane (http://cochrane.org) is an international non-profit organization dedicated to helping healthcare professionals, policy makers and patients make healthcare decisions. Cochrane, with more than 50 different review groups, focuses on creating and publishing Cochrane systematic reviews of very high quality. Cochrane systematic reviews (e.g. protocols, reviews and review updates) are published in The Cochrane Database of Systematic Reviews in The Cochrane Library (http://www.cochranelibrary.com/). Unlike other journals, Cochrane systematic reviews must be periodically updated. The reason for these updates is that new studies may change the conclusions of the systematic review. This ensures that decisions about pediatric oncology care can be made based on the most up-to-date evidence.

Cochrane encourages the conduct of systematic reviews on interventions and diagnosis for cancer in children with respect to prevention, treatment, diagnosis, supportive care, psychosocial care, palliative and terminal care, nursing care and late effects of treatment. Within the Princess Máxima Center for Pediatric Oncology, the Systematic Review & Guideline Unit provides support for these Cochrane systematic reviews.

1.4 Structure of the handbook

In this handbook information is provided that will be useful to systematic review author teams as they develop their review. Specifically it will:

1. Outline the key steps in the development of systematic reviews.
2. Direct review authors to other important and more detailed sources of information integral to systematic review development.
3. Provide practical information regarding the organisation of the systematic review development.

2 Methodology used by the Systematic Review & Guideline Unit to develop systematic reviews

Conducting a systematic review is a structured process and consists of three phases:
1. Preparation phase
2. Development phase
3. Finalisation phase

2.1 Preparation phase

Avoid duplicate work

Duplicate work, i.e. starting another review looking at the same topic, is of course a waste of resources. So before you start working on your review always check if such a review has not yet been performed. You should check publications in for example PubMed/MEDLINE or registered protocol registries (for more information see point 2.2 below).

Review author team

Convening a review author team is a crucial step in developing a systematic review. Diversity is an essential feature of this team. Its exact composition should be tailored to the review topic and reflect the range of stakeholders involved. The team should at least include clinical content experts and systematic review methodology experts, but it is preferable to also include patients or their representatives.

2.2 Development phase

In general, the systematic review development process consists of seven key steps:
1. Formulate a clinical question
2. Develop a comprehensive search strategy to identify all relevant evidence
3. Select relevant studies on the basis of clear predefined inclusion and exclusion criteria
4. Extract data from the included studies
5. Assess the risk of bias in the included studies
6. Planning the analyses
7. Preparing Summary of Findings tables, potentially including a GRADE (Grading of Recommendations Assessment, Development and Evaluation) assessment

When you have prepared the protocol for the review you should consider to prospectively register it in order to avoid duplicate work and to be transparent about the planned methodology so that it can be compared what was actually done. Cochrane systematic review protocols are automatically published in the Cochrane Library. Other reviews can for example be registered in PROSPERO (https://www.crd.york.ac.uk/PROSPERO/).

Step 1: Formulate a clinical question

The clinical question should be clear and relevant to clinical practice. Formulating the question can be done using the PICO (Participants, Interventions, Comparisons, Outcomes) model:

- What are the characteristics of the participants you are interested in? (e.g. type of malignancy, stage of disease, gender, age, minimal survival time)
- What is the intervention you want to evaluate? (e.g. type of treatment)
- What is the comparison you are interested in? (e.g. different type of treatment, no intervention; this is not always relevant)
- What are relevant outcomes? (e.g. survival, adverse effects)

Examples of the PICO method and the formulation of a clinical question are shown in Table 1.

Table 1. Examples of the PICO method and clinical questions

Example 1	
P (Participants)	Children with cancer
I (Intervention*)	Platinum-based therapy including a medical intervention to prevent hearing loss
C (Comparison**)	Platinum-based therapy including placebo, no additional treatment or another protective medical intervention
O (Outcomes)	Hearing loss, anti-tumour efficacy, toxicities other than hearing loss and quality of life
Clinical question	What is the efficacy of medical interventions to prevent hearing loss and to determine possible effects of these interventions on anti-tumour efficacy, toxicities other than hearing loss and quality of life in children with cancer treated with platinum-based therapy as compared to placebo, no additional treatment or another protective medical intervention?
Example 2	
P (Participants)	Children with cancer
I (Intervention*)	Platinum-based therapy
C (Comparison**)	Not applicable
O (Outcomes)	Hearing loss
Clinical question	What is the association between childhood cancer treatment including platinum analogues and the occurrence of hearing loss?

** In diagnostic reviews the I stands for Index test, in prognostic reviews the I stands for the presence of a certain prognostic factor; ** In diagnostic reviews the C stands for Comparator test, in prognostic reviews the C stands for the absence of the prognostic factor included in I.*

Step 2: Develop a comprehensive search strategy to identify all relevant evidence

It is important to identify all available evidence eligible for inclusion in the review, both published and unpublished (within the limits of resources and time). A suboptimal literature search can compromise the validity of the conclusions of the systematic review as it can introduce reporting bias.

Statistically significant results that indicate that an intervention works are more likely to be published, more likely to be published rapidly, more likely to be published in English, more likely to be published more than once, more likely to be published in high impact journals and more likely to be cited by others than other results. When the dissemination of study findings is influenced by the nature and direction of its results reporting bias can occur. There are different types of reporting bias: publication bias, time lag bias, duplicate publication bias, location bias, citation bias, language bias and outcome reporting bias [1].

To avoid reporting bias the search for the systematic review needs to be as comprehensive as possible, using several different sources and a sensitive search strategy.

Sources to consider are:

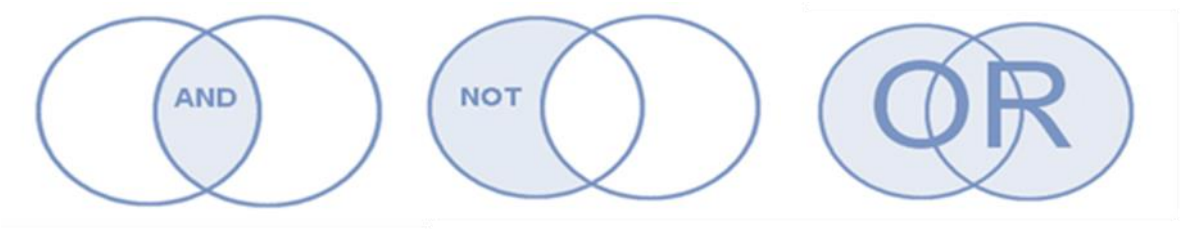
- Different electronic bibliographic databases, like PubMed/MEDLINE, Embase and the Cochrane Central Register of Controlled Trials (CENTRAL). There are also specialty databases like PsycInfo (psychology topics) and CINAHL (nursing topics).
  - Although there is some overlap between these databases, differences do exist, so it is important to search several of them.
- Reference lists of relevant articles (included studies, other reviews and key papers)
- Conference proceedings, like SIOP (Société Internationale d'Oncologie Pédiatrique/International Society of Paediatric Oncology) and ASPHO (American Society of Pediatric Hematology/Oncology).
  - For example the last 2 or 5 years can be checked.
- Ongoing trial registries, like the World Health Organization International Clinical Trials Registry Platform (WHO ICTRP) or Clinicaltrials.gov.
- Experts in the field (to identify studies not found in other sources)

For most of these sources a specific search strategy must be developed. It is strongly advised that this is done in collaboration with a trained information specialist.

The clinical question serves as the starting point for the development of a sensitive search strategy. The search strategy is based on the PICO model and should use controlled vocabulary (which is specific for each database, like MeSH in PubMed/MEDLINE and EMTREE in Embase) and free-text words. It is important to consider all of the related terms, variations in spellings and synonyms for each PICO item included in the search.

Different search terms can be combined using Boolean operators like "OR" (retrieves articles labelled by at least one of the search terms), "AND" (retrieves only articles labelled by all search terms), and "NOT" (excludes search terms from the search) (see Figure 1). "NOT" must be used very carefully as important publications may be missed when it is included in the search strategy. An example of a full search can be found in Appendix 1.

Figure 1. Use of Boolean operators



Some validated search strategies and filters are available, for example for children [2] and randomized controlled trials and controlled clinical trials [1]. It is advised to use these validated strategies whenever relevant for your review.

In order to be transparent and reproducible it is important to report your search strategy in detail: which database, what interface, complete and detailed search terms, the date the search was run in each source, the years covered by the search and, if applicable, use of any limits (see step 3 for more information).

Standard search strategies developed by the Systematic Review & Guideline Unit are presented in the Handbook for Guideline Development (for the latest version see: https://www.ighg.org/international-guideline-harmonization-group/methods/handbook/). Also, all published Cochrane protocols and reviews on childhood cancer can be found at the Cochrane Database of Systematic Reviews in The Cochrane Library (http://www.cochranelibrary.com/). These include the search strategies used for different databases. You can check if a search relevant to your review topic already exists. However, as these search strategies were developed over the course of several years and some databases have implemented changes during this time period, please be aware that they might need to be adjusted for your current search to be reliable.

More information on developing a search strategy:

- ✓ *Lundh A, Kremer LCM, Leclercq E. Development of a search strategy. Evid.-Based Child Health 2007; 2(2): 937-939*
- ✓ *Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from https://handbook-5-1.cochrane.org/*
- ✓ *Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook*
- ✓ *Van Dalen EC, Kremer LC. Inclusion of studies into a Cochrane review. Evid.-Based Child Health 2006; 1(4): 1349-1351*

Step 3: Select relevant studies on the basis of clear predefined inclusion and exclusion criteria

Inclusion and exclusion criteria

The selection of studies for inclusion in the review should be based on clearly defined in- and exclusion criteria based on the PICO model. The following points should be considered, but for each specific review question additional issues can apply:

- Participants:
  - Age, including the time-point when this should be assessed (e.g. at diagnosis, at treatment, at follow-up)
  - Gender
  - Disease, for example childhood cancer diagnosis, stage of disease, newly diagnosed or relapsed patients
  - Time since cancer diagnosis (e.g. during treatment, X-year survivors after diagnosis, X-year survivors after end of therapy)
- Intervention and comparison treatments: be very specific (for example, should only one chemotherapeutic agent be different between the treatment groups, dosage, timing)
- Outcomes; for each outcome decide if you are:
  - Only interested in a specific definition or in all definitions used by the included studies (for example, only adverse effects defined based on National Cancer Institute Common Terminology Criteria for Adverse Events version 3 (NCI-CTCAEv3) criteria or also other versions and other criteria like the World Health Organization (WHO) criteria)
  - Only interested in multivariable risk factor analyses or also univariable ones (when you are looking at risk factors in your review)
  - Be aware that if a study does not report any usable outcome data this is not a reason for exclusion

Other considerations:

- Study design: choose the most optimal study design to answer your specific review question. For example randomized controlled trials for intervention questions or diagnostic test accuracy studies for diagnostic questions.
- When you are addressing the occurrence of adverse effects after a certain treatment in your review you might encounter randomized controlled trials or controlled clinical trials in which participants in both groups received the treatment you are interested in. In that case you might be able to include them as an observational study in your review.
- When a cross-over trial is available in which the second period does not reliably answers your question (for example because of a carry-over effect), you might still be able to use only information from the first part.
- Language: ideally, in order to avoid language bias, there should not be a language restriction.
- Specific search dates, i.e. published from a specific date onwards: this is only recommended when you can be certain that important studies won't be missed, for example when a diagnostic test was not yet available for the participants you are interested in.
- Minimal sample size: for example at least 50 participants depending on the clinical problem and availability of evidence. However, these cut-offs are arbitrary and you should be aware that important information can be missed.

In addition to clearly defining the in- and exclusion criteria you should decide what will be done when only a part of study participants are eligible for inclusion in your review. For example, when you are interested in participants aged 0-18 years at cancer diagnosis, but also participants diagnosed when older than 18 years are included. Or when not only malignant tumors are included, but also some benign conditions. Options are to only include data on eligible participants (but then this information should be available) or to define a cut-off value for the minimal percentage of participants that fulfils the inclusion criteria for the review.

Finally, you should predefine what you will do when there is overlap between some of the studies eligible for the review as sometimes studies are published more than once or the same participants are included in more than one publication. Options are to only include one study (for example the most recent one or the one that provided the most information) or to combine information from different studies.

Study selection

When the searches have been run and the results are de-duplicated (for example in endnote software: https://endnote.com/) two independent reviewers should assess if publications meet the inclusion criteria. This reduces the possibility of missing relevant papers [3]. This will not be feasible for all search sources. For example cross-checking reference lists of relevant papers can initially be done by one reviewer, but when a possible relevant study is identified a second independent reviewer should confirm its eligibility for inclusion in the review.

There are different phases in the study selection process:

1. Title and/or abstract phase:
   In the title/abstract phase you can choose: eligible, not eligible, or unclear if eligible. In case of doubt always select the article for fulltext assessment. The results of the reviewers' assessments should be compared and discrepancies discussed and resolved. If that is not possible, a so-called third party arbitration by another reviewer can help.
2. Obtain fulltext publications for articles labelled as eligible or unclear if eligible for inclusion in the review. Some will be publicly available and others will be available at your local library (or can be requested there). If not you can for example request it from one of the authors of the publication
3. Fulltext phase:
   In the fulltext phase you can choose: eligible or not eligible. Again, the results of the reviewers' assessments should be compared and discrepancies discussed and resolved. If that is not possible, a so-called third party arbitration by another reviewer can help. Reasons for study exclusion should be noted.

There is software available that can help you with this step of the review process, for example Rayyan (https://rayyan.qcri.org/welcome).

Reviews, either systematic or narrative, are not eligible for inclusion, but reviewers should screen reference lists of important reviews to obtain relevant papers not included in other search sources. Reviewers can for example make a list of possible relevant reviews during the title and/or abstract phase and obtain the fulltext publications to look at the reference lists.

When during study selection you encounter missing information needed to make a decision for inclusion or exclusion, then try to obtain the necessary data from the authors of the study.

Before the final inclusion in the review always check errata and retraction statements of the eligible studies.

More information on study selection:

- ✓ *Van Dalen EC, Kremer LC. Inclusion of studies into a Cochrane review. Evid.-Based Child Health 2006; 1(4): 1349-1351*
- ✓ *Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from https://handbook-5-1.cochrane.org/*
- ✓ *Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook*

Step 4: Extract data from the included studies

Ideally two independent review authors should extract data for each included study on at least study design, participants, interventions, outcomes, risk of bias (see step 5 below) and follow-up using a data extraction form. However, if that is not possible (within the limits of resources and time) it should at least be done by one review author and checked by another one. The results of the reviewers' assessments should be compared and discrepancies discussed and resolved. If that is not possible, a so-called third party arbitration by another reviewer can help. Always make sure that reviewers who are a co-author of one of the included studies are not involved in the data extraction process of that study.

The format of the data extraction form can be like an evidence table, but the exact presentation format can depend on journal requirements. An example of an evidence table is included in appendix 2. It is useful to prepare instructions on how to use the form and to pilot test the form in order to avoid different interpretations by different authors performing or checking the data extraction.

For general information you should preferably focus on information for all patients eligible for the review, not only on patients with an outcome assessment. If that information is not provided you can consider to report the available information, but with a clear statement what you did.

For treatment dose, it will be most informative to report information on the actual received cumulative dose. Many studies report only the dose according to protocol without reporting dose adjustments due to for example toxicity; it should always be clear what you report.

When during data extraction you encounter missing information try to obtain the necessary data from the authors of the study.

Step 5: Assess the risk of bias in the included studies

Ideally two independent review authors should perform the risk of bias assessment, but if that is not possible (within the limits of resources and time) it should at least be done by one review author and checked by another one. The results of the reviewers' assessments should be compared and discrepancies discussed and resolved. If that is not possible, a so-called third party arbitration by another reviewer can help. Always make sure that reviewers who are a co-author of one of the included studies are not involved in the risk of bias assessment of that study.

Different checklists are available to assess the risk of bias in included studies and largely depend on the type of question the review is trying to answer. See table 2 for examples. Always check if these criteria need to be further specified for your review.

Table 2. Examples of checklists to assess risk of bias for different types of eligible studies

Type of eligible studies	Checklist example	More information
(Randomized) controlled trials addressing a therapeutic question; parallel design*	• Cochrane Risk of Bias version 1 • Cochrane Risk of Bias version 2	For version 1: • Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from https://handbook-5-1.cochrane.org/ For version 2: • Online version: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook • Paper version: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions. 2nd Edition. Chichester (UK): John Wiley & Sons, 2019.
Observational studies addressing adverse effects	Cochrane Childhood Cancer risk of bias assessment criteria for observational studies	See Appendix 3
Diagnostic test accuracy studies	QUADAS-2	https://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/
Prognosis/prediction model studies	PROBAST	https://methods.cochrane.org/prognosis/tools
Qualitative studies	CASP checklist for qualitative studies	https://casp-uk.net/wp-content/uploads/2018/01/CASP-Qualitative-Checklist-2018.pdf

**when cluster-randomized trials or cross-over trials are included some additional issues, like the carry-over effect, need to be taken into account when performing the risk of bias assessment*

In general there are 3 options when judging the risk of bias in a study: low risk of bias, high risk of bias and unclear risk of bias. Some risk of bias items should be assessed for each outcome separately, for example attrition bias and detection bias (the latter one with the exception of overall survival as this is an objective outcome for which the risk of detection bias is negligible). It is useful to prepare instructions on how to use the risk of bias checklist and to pilot the assessment in order to avoid different interpretations by different authors.

When during risk of bias assessment you encounter missing information try to obtain the necessary data from the authors of the study.

More information on risk of bias assessment:

- ✓ *Van Dalen EC, Kremer LCM, Moyer VA. Quality of studies included in a systematic review and associated risk of bias - garbage in, garbage out. Evid.-Based Child Health 2007; 2(4):1321-1324*
- ✓ *Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from https://handbook-5-1.cochrane.org/*
- ✓ *Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook*

Step 6: Planning the analyses

Issues to consider when planning the analyses of your systematic review:

a. Types of data, effect measures and unit of analysis
b. Meta-analysis and heterogeneity
c. Missing data
d. Subgroup analyses
e. Sensitivity analyses
f. Detecting the presence of reporting bias
g. Other considerations

It is strongly advised that (meta-)analyses are done in collaboration with a trained epidemiologist or statistician.

Different software packages for systematic reviews are available. Examples are Review Manager (RevMan) for Cochrane systematic reviews or the systematic review packages of R.
Not all software allows you to perform all analyses.

a. Types of data, effect measures and unit of analysis

Types of data and effect measures

First you need to decide what the most appropriate type of data for your specific review question is, keeping in mind that sometimes an outcome can be described using different data types.

For example, cardiac function can be addressed in a dichotomous or binary fashion (i.e. how many participants had an abnormal cardiac function test?) or in a continuous fashion (i.e. what was the mean value of the cardiac function test in the different treatment groups?). When you use the latter option it is possible that patients with good and bad values on their cardiac function test balance each other out resulting in an adequate mean value. This can give the impression that there is no problem, while for some patients this might not be true. So it can be more informative to analyse this outcome as dichotomous.

Another example are time-to-event or survival data. When the status of all study participants at a fixed time-point (e.g. 5-year survival) is known they can sometimes be analyzed as dichotomous data. However, it should be kept in mind that bias can arise if the time-points are subjectively chosen (either by the review authors or the authors of the included study). The most appropriate way of summarizing time-to-event data is to present the treatment effect as a hazard ratio. When not all necessary data are available Parmar's methods can be used to try to obtain the missing data [4].

Ordinal outcomes can be addressed as continuous data or they can be dichotomized using a cut-off point (like pain above or below the level of requiring an intervention).

For each type of data different effect measures can be calculated (see Table 3).

Table 3. Examples of effect measures for different types of data

Type of data	Examples of effect measures
Dichotomous/binary Ordinal (when analysed as dichotomous)	Relative risk/risk ratio Relative risk reduction Absolute risk reduction/risk difference Number needed to treat or harm Odds ratio Prevalence (Cumulative) incidence
Continuous Ordinal (when analysed as continuous)	Mean difference (when all studies use the same scale) Standardized mean difference (when studies use different scales)
Time-to-event	Hazard ratio
Diagnostic test accuracy	Sensitivity Specificity Positive predictive value Negative predictive value

More information on types of outcomes and effect measures:

- ✓ *Kremer LCM, Moyer V. Tips and tricks for understanding and using SR results. Evid.-Based Child Health 2006; 1(1):356-358*
- ✓ *Kremer LCM, Barrowman N. Odds and odds ratio. Evid.-Based Child Health 2006; 1(2):732-733*
- ✓ *Van Dalen EC, Tierney JF, Kremer LC. Time-to-event data. Evid-Based Child Health 2007; 2(3): 1089-1090*
- ✓ *Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 1998;17(24):2815-2834. Erratum in: Stat Med 2004;23(11):1817*
- ✓ *Van der Lee JH. Outcomes - don't be misled by results in which you are not interested Evid.-Based Child Health 2008;3(4):1153-1155*
- ✓ *Leeflang MMG, Bossuyt PMM. Systematic reviews of diagnostic test accuracy. Evid.-Based Child Health 2008; 3(1): 257-258*
- ✓ *Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from https://handbook-5-1.cochrane.org/*
- ✓ *Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook*

Unit of analysis

Most often the unit of analysis will be a single measurement for each outcome in the individual study participants. However, when for example cluster-randomized trials or cross-over trials are eligible for inclusion in the review this needs to be taken into account.

The same is true when there are multiple measurements for the same outcome, like with repeated measurements (for example at the end of treatment and 1 year after the end of treatment). You should thus always define the time points for outcome measurements you want to address in the review.
Other unit of analysis issues can arise with re-occuring events, multiple treatment attempts, multiple body parts and more than two intervention groups.

More information on unit of analysis:

- ✓ *Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from https://handbook-5-1.cochrane.org/*
- ✓ *Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook*

b. Meta-analysis and heterogeneity [5]

A meta-analysis can, but doesn't have to be, one of the components of a systematic review. In a meta-analysis the results of several individual studies are combined and the presence of relatively small effects can be more easily detected. However, a meta-analysis should only be performed if the study design, participants, interventions, and outcomes (including outcome definitions) in the individual studies are comparable.

Inevitably, studies brought together in systematic reviews will differ. Variability among the individual studies is called heterogeneity. There are different types of heterogeneity:

- clinical heterogeneity (caused by variability in the participants, interventions and outcomes studied)
- methodological heterogeneity (caused by variability in study design and quality)

A consequence of clinical and/or methodological heterogeneity is the occurrence of statistical heterogeneity, which is variability in the treatment effects being evaluated in the different studies. Statistical heterogeneity manifests itself in the observed treatment effects being more different from each other than one would expect due to chance alone.

If it is already clear that studies differ too much when looking at important characteristics including used outcome definitions, then a meta-analysis should not be performed at all and results should be presented descriptively. However, not all issues that can cause heterogeneity will be known in advance. So when you have performed a meta-analysis you should always check for the presence of heterogeneity.

There are different methods to explore the presence of heterogeneity, for example:

- Visual inspection of the graphical display of the results: if the confidence intervals for the results of the individual studies have poor overlap, this generally indicates the presence of heterogeneity.
- The I^2 statistic: this describes the percentage of the total variability in effect estimates that is due to heterogeneity rather than within study variation. A value of 0% indicates no observed heterogeneity, and larger values show increasing heterogeneity. A value above 50% is often considered substantial heterogeneity.

If heterogeneity is identified there are several options as how to proceed:

1. Check if the data are correct:
   For example, errors like mistakenly entering standard errors as standard deviations may cause apparent heterogeneity.
2. Do not perform a meta-analysis:
   Particularly if there is inconsistency in the direction of effect or if only two or three studies are available that differ largely in their results, it may be misleading to quote an average value for the treatment effect.
3. Explore heterogeneity:
   Look for apparent differences between studies in more detail when deciding to perform the meta-analysis in the first place. Subgroup meta-analyses or, more sophisticated and complicated, meta-regression can be conducted.
4. Ignore heterogeneity by performing a fixed effects meta-analysis*:
   However, as the assumptions of a fixed effects model imply that the observed differences among individual study results are solely due to chance, i.e. that there is no heterogeneity, this is not a good strategy and results can be misleading.
5. Perform a random effects meta-analysis*:
   This may be used to incorporate heterogeneity among individual studies and is intended for heterogeneity which cannot be explained. But keep in mind that when using a random effects model the presence of heterogeneity is still an issue.
6. Change the effect measure:
   The choice of the effect measure (like odds ratio or relative risk) may affect the degree of heterogeneity among results.
7. Exclude studies:
   This may introduce bias, but if an obvious reason for an outlying result of a study is apparent the study might be excluded from the meta-analysis with more confidence.

** a substantial difference in the overall effect estimate (and corresponding confidence interval) calculated by the fixed and random effects models will be seen only if studies are markedly heterogeneous.*

When performing a meta-analysis, always check if you don't include the same patients twice (for example in case of overlapping publications).

Also, be careful with including results of trials that are only available as a conference abstract in your analyses (instead of presenting them separately in a descriptive manner) as results often differ significantly between both publications [6].

More information on meta-analysis and heterogeneity:

- ✓ *Kremer LCM, van Dalen EC, Vandermeer B, Offringa M. Meta-analysis and heterogeneity. Evid.-Based Child Health 2010;5(1):12-16*
- ✓ *Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from https://handbook-5-1.cochrane.org/*

- ✓ *Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook*

c. Missing data

If possible, intention-to-treat analyses should be performed; if this is not possible, an explanation should be provided.

Consider if data are "missing at random" (i.e. unrelated to the actual value of the missing data) or "not missing at random" (i.e. related to the actual value of the missing data).

The principal options for dealing with missing data are [1]:

1. Analysing only the available data (i.e. ignoring the missing data).
2. Imputing the missing data with replacement values, and treating these as if they were observed (e.g. last observation carried forward, imputing an assumed outcome such as assuming all were poor outcomes (worst-case) or all were good outcomes (best-case), imputing the mean, imputing based on predicted values from a regression analysis).
3. Imputing the missing data and accounting for the fact that these were imputed with uncertainty (e.g. multiple imputation, simple imputation methods (as point 2) with adjustment to the standard error).
4. Using statistical models to allow for missing data, making assumptions about their relationships with the available data.

Be aware that usually best-case means assuming that a participant does not have the outcome (for example heart failure). However, for example for tumour response (i.e. number of patients with a remission) this is the opposite: due to the nature of this outcome best-case here means that the participant does have the outcome.

It is recommended to involve a statistician when imputing data using more advanced methods (i.e. other than a best-case and worst-case analysis).

More information on missing data:

- ✓ *Hooton N, Sumamo E, Liang Y. Intention to treat analysis. Evid.-Based Child Health 2008;3(2):591-592*
- ✓ *Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from https://handbook-5-1.cochrane.org/*
- ✓ *Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook*

d. Subgroup analyses [7]

Subgroup analyses may be performed for subsets of participants (based on for example gender, age or severity of disease), for subsets of intervention characteristics (for example dose and duration of treatment), or for subsets of studies (like different geographical locations).

The decision to perform a subgroup analysis should always be based on a sound scientific rationale (i.e. biological, sociological or clinical hypotheses). It should ideally be supported by evidence from sources other than the studies included in the review.

Subgroup analyses should preferably be planned in advance (i.e. a priori). However, it is possible that during the review process it becomes clear that a particular subgroup is important. Then the subgroup analysis can be performed, but it should be made clear in the review that it was a post hoc analysis and its results should be treated with caution.

Without a substantial number of studies (i.e. with a low statistical power), it is highly unlikely that a subgroup analysis will give useful results. The following can be used as a rule of thumb: at least ten studies should be available for each evaluated subgroup. However, even this number can be too few. As a result, false-negative results cannot be ruled out.

You should only perform a limited number of subgroup analyses as performing large numbers of analyses will always lead to some significant results. By definition, testing at the 5% level of significance will erroneously report a statistically significant difference between subgroups in about 5% of the performed analyses (i.e. false-positive results).

Also, keep the possibility of confounding in mind.

More information on subgroup analyses:

- ✓ *Van Dalen EC, Kremer LCM. Subgroup analyses. Evid.-Based Child Health 2009; 4(2):1140-1141*
- ✓ *Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from https://handbook-5-1.cochrane.org/*
- ✓ *Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook*

e. Sensitivity analyses

Performing a systematic review involves many choices, some of which may be subjective or unclear. For example an age limit in the inclusion criteria, different statistical methods available to do the analyses and no consensus on what the best method is, missing data for the risk of bias assessment or outcome assessments, etcetera.

To prove that the results of the systematic review are not dependent on subjective or unclear choices, sensitivity analyses can be performed. This is a repeat of the meta-analysis taking into account the subjective or unclear choices.

For example, if trials with a high or unclear risk of bias and trials with a low risk of bias are included in a meta-analysis, a sensitivity analysis can be performed to explore whether trial quality plays a role in determining the effect size. Studies with a high risk of bias and studies for which the risk of bias is unclear need to be excluded, and the results of only studies with a low risk of bias should be compared with the results of all available studies. This should be done for all risk of bias items separately.

It is advised to only perform sensitivity analyses if at least two studies remain in the analysis after exclusion of some of the studies.

More information on sensitivity analyses:

- ✓ *Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from https://handbook-5-1.cochrane.org/*
- ✓ *Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook*

f. Detecting the presence of reporting bias

Unfortunately, performing comprehensive searches doesn't guarantee that there is no reporting bias present in your review. By preparing a so called funnel plot it is possible to investigate whether a review is subject to reporting bias. As a rule of thumb, tests for funnel plot asymmetry should be used only when there are at least 10 studies included in the meta-analysis, because when there are fewer studies the power of the tests is too low to distinguish chance from real asymmetry [1]. Several methods can be used to test for asymmetry in a funnel plot; it is recommended to involve a statistician when doing this.

It should however be kept in mind that while an asymmetric funnel plot can be the result of reporting bias, other reasons should also be considered. Vice versa a symmetrical funnel plot does not necessarily rule out the presence of reporting bias.

More information on reporting bias:

- ✓ *Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from https://handbook-5-1.cochrane.org/*
- ✓ *Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook*

g. Other considerations

When for a particular outcome only one study is available and there are no events in one of the treatment groups, it is impossible to calculate a relative risk or odds ratio. Some statistical packages get round this by adding half a case to the treatment group with no events. If you are doing a meta-analysis with many studies and most of these studies have events in both treatment groups, adding an extra half event in one treatment group doesn't make much difference to the overall estimate of the relative risk or odds ratio. However, if you have only one study and you add half an event to one treatment group, the relative risk or odds ratio, its 95% CI and the p-value can be misleading. For these outcomes you should calculate the Fischer's exact p instead.

When no events are observed a confidence interval can be calculated using the so called "rule of three" [8].

Step 7: Preparing Summary of Findings tables, potentially including a GRADE (Grading of Recommendations Assessment, Development and Evaluation) assessment

The final step of the development phase is the preparation of Summary of Findings tables. They provide an overview of the body of evidence for a specific clinical question and include information on study and patient characteristics, outcomes and associated effect estimates (either the overall pooled result, or if pooling was not possible results for each individual study separately).

Authors can develop their own Summary of Findings table or they can be prepared using GRADE profiler software (https://gradepro.org/). An example of a Summary of Findings table for Cochrane systematic reviews is shown in appendix 4. An example of a Summary of Findings table for observational studies is shown in appendix 5.

When comparative risks have been calculated check if they are in line with the identified effect estimate.

If possible the Summary of Findings table should also include an assessment of and rationale for the quality of the evidence using the GRADE methodology (https://www.gradeworkinggroup.org/). The quality of a body of evidence is defined as the extent to which one can be confident that an identified effect or association is true.

Ideally two independent review authors should perform the GRADE assessment, but if that is not possible (within the limits of resources and time) it should at least be done by one review author and checked by another one. The results of the reviewers' assessments should be compared and discrepancies discussed and resolved. If that is not possible, a so-called third party arbitration by another reviewer can help.

The evidence is graded according to four levels (https://gdt.gradepro.org/; assessed 2 March 2021):

- 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.

The initial score (start point) is based on study type and review question. For example for intervention reviews randomized studies start at high level of evidence, while controlled clinical trials and observational studies start at a low level of evidence.

There are 5 reasons for downgrading the level of evidence:

1. Study limitations (risk of bias)
2. Inconsistency of the results (like a wide variance in point estimates and minimal or no overlap in 95% confidence intervals or unexplained heterogeneity)
3. Indirectness of the study population, interventions and outcomes
4. Imprecision of the effect estimates (as a rule of thumb when there are less than 300 events you should consider downgrading by at least 1 level [9].
5. Publication bias

There are 3 reasons to upgrade the level of evidence:

1. Large magnitude of effect
2. Dose response gradient
3. Plausible confounding

Be aware that not every study can be upgraded. The possibility to upgrade the level of evidence depends on the results of the 5 reasons for downgrading and the study design.

The reasons to downgrade or upgrade the level of evidence are explained in detail in the GRADEpro handbook (https://gdt.gradepro.org/app/handbook/handbook.html) and in the different publications of the GRADE working group, especially in the Journal Of Clinical Epidemiology (https://www.gradeworkinggroup.org/; be aware that the list provided here is not up-to-date, more publications are available). GRADE is somewhat subjective, so review authors should always report their decisions in a transparent manner.

More information on Summary of Findings tables and GRADE assessment:

- ✓ *Langendam MW, Kuijpers T, de Beer H, Kremer LCM. An introduction to the GRADE approach; rating the quality of evidence for an intervention Evid.-Based Child Health 2010; 5(2):537-540*
- ✓ *Schünemann H, Brożek J, Guyatt G, Oxman A (editors). Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Available from https://gdt.gradepro.org/app/handbook/handbook.html*
- ✓ *Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from https://handbook-5-1.cochrane.org/*
- ✓ *Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook*
- ✓ *Mulder RL, Brown MC, Skinner R, Hudson MM, Kremer LCM. Handbook for guideline development; collaboration between International Guideline Harmonization Group, PanCare Guideline Group and Cochrane Childhood Cancer Group 2019. Available from www.ighg.org*
- ✓ *Skoetz N, Goldkuhle M, van Dalen EC, et al. GRADE guidelines 27: how to calculate absolute effects for time-to-event outcomes in summary of findings tables and Evidence Profiles. J Clin Epidemiol. 2020;118:124-131*
- ✓ *Other articles from the GRADE working group in the Journal of Clinical Oncology*

2.3 Finalisation phase

Writing the review

All systematic reviews should be summarized in a manuscript appropriate for publication in a peer-reviewed journal. You can use for example the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses" (PRISMA) checklist (http://www.prisma-statement.org/) when preparing the manuscript, but always check if the journal has other requirements or advises another checklist.

The manuscript should include at least the following items:

- Title
- Abstract
- Background including the rationale for performing the review
- Objective(s)/clinical question(s)
- Methods:
  - inclusion and exclusion criteria
  - search strategy
  - study selection
  - data extraction
  - risk of bias assessment of included studies
  - analysis plan
  - GRADE assessment of included studies (if included in the review)
- Results:
  - results of the search
  - characteristics of included studies
  - description of the evidence
  - risk of bias assessment
  - GRADE assessments (if performed)
- Discussion:
  - short summary of results
  - applicability of the evidence to clinical practice
  - strengths and limitations of the review
  - conclusions
  - implications for future research (if needed)
- Reference list
- Relevant tables and figures

The exact information that needs to be reported for each item differs with each review, so the list below is not exhaustive, but things to keep in mind when preparing your manuscript are:

General

- Make sure that results are presented the same throughout the manuscript. When you change something at one location make sure to do this at all other relevant locations. This might seem obvious, but often goes wrong.
- All efforts made to obtain additional data from authors of included studies or trials considered for inclusion should be reported and it should be stated which data it does concern.

Title

This should include either systematic review, meta-analysis or both.

Objective(s)/clinical question(s) and methods

Provide a detailed description of the methodology and definitions you used in your review. See the information provided at the development phase for important issues to include.

Results

Results of the search

Make sure to include information on all different sources including search dates and number of identified titles/abstracts for each (when relevant).

If relevant, report (possible) overlap between studies and how this was handled in the review.

The results of the search should be presented in a flow diagram of the selection of studies. It should show the number of records identified for each search source, the number of records after deduplication, the number of records included and excluded during title and abstract screening, and the number of papers included and excluded during fulltext assessment (including the reasons for exclusions). A template for such a flow diagram can be found at http://www.prisma-statement.org/.

Characteristics of included studies

Provide a short summary of the main characteristics of the included studies. The exact information depends on the individual review, but information should be provided on for example study design of included studies, total number of patients included in the included studies, total number of patients in intervention and control groups, a short description of the interventions, age, type(s) of malignancies, length of follow-up. Refer readers to the relevant tables for more detailed information.

Description of the evidence

Report for each outcome/analysis (as sometimes there is more than one analysis for a certain outcome):

- How many studies (with references) did add to the body of evidence
- How many participants did add to the body of evidence
- What type of analysis has been performed (if not an intention-to-treat analysis describe the reasons)
- The effect estimate and accompanying 95% confidence interval, P-value, and if pooled results also the results of the I^2-test
- If there was a significant difference between treatment groups (and if so: in favour of which group)
- If relevant, also number of patients with an event (for dichotomous data) or means (for continuous data) in the different treatment groups
- If relevant, the outcome definitions as used in the included studies and review
- When for time-to-event data Parmar's method was used to obtain necessary information this should be reported

- If relevant, the follow-up duration(s)
- When comparative risk have been calculated the basis for the assumed risk should always be reported.

Risk of bias assessment

Give a short summary of the identified risk of bias in the included studies; for each item preferably all studies should be reported (i.e. n=… (…%) low risk of bias, n=… (…%) high risk of bias and n=… (…%) unclear risk of bias).

The judgements for each risk of bias item for each study/outcome including the exact reasoning can be reported in a table. You can also consider including a risk of bias figure.

Usually an overall score should not be reported, but always check the risk of bias assessment tool you have used.

When you addressed selective reporting make sure that it is clear how this was done, either in the methods or in the results section). Did you for example only compare the methods section of an included manuscript with the results section or did you look in a more extensive manner.

An example of a Risk of bias summary figure can be found in figure 2. An example of a Risk of bias support for judgement table can be found in table 4. Please note that in Cochrane systematic reviews, where this example comes from, these tables include only one included study and one review can thus include several of these tables. You can of course format the table to include all included studies.

Figure 2. Example of a risk of bias summary figure [10]

Table 4. Example of a Risk of bias support for judgement table [10; Brock 2018]

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Stated that participants were randomly assigned to sodium thiosulfate or no additional treatment, but no further information on the methods of randomization provided.
Allocation concealment (selection bias)	Unclear risk	Stated that participants were randomly assigned to sodium thiosulfate or no additional treatment, but no further information on the methods of randomization provided.
Blinding of participants and personnel (performance bias)	Unclear risk	No information on blinding of participants and personnel provided.
Blinding of outcome assessment (detection bias) - ototoxicity	Low risk	For hearing loss, outcome assessors were blinded (quote: "The trial data were blinded for the audiology central reviewer")
Blinding of outcome assessment (detection bias) - overall survival	Low risk	No information on blinding of outcome assessors provided, but as blinding is not relevant for the outcome overall survival we judged this outcome at low risk of detection bias.
Blinding of outcome assessment (detection bias) - other reported outcomes	Unclear risk	No information on blinding of outcome assessors provided for event-free survival, response rate and adverse effects other than ototoxicity.
Incomplete outcome data (attrition bias) - ototoxicity	Low risk	Hearing loss evaluated in 101/109 children (in the sodium thiosulfate group, 1 child died before the definitive hearing test and 1 child was not tested; in the control group, 4 children died before the definitive hearing test and 2 children were not tested; the reasons 3 children were not tested were audiometry not feasible for health reasons (2 children) and parent refusal (1 child)).
Incomplete outcome data (attrition bias) - tumour response and adverse events	Low risk	For tumour response, 1/52 participants in the control group was not evaluated (reason nm); based on the provided information, we assumed all participants were evaluated for adverse effects.
Incomplete outcome data (attrition bias) - survival (overall, event-free or both)	Unclear risk	Insufficient information provided to adequately judge this outcome.
Selective reporting (reporting bias)	Low risk	A protocol was mentioned in the manuscript and all expected outcomes were reported.
Other bias	Unclear risk	*Block randomization in unblinded trials*: unclear (information on both method of randomization and blinding of participants and care providers and for most outcomes blinding of outcome assessors was not provided).

		Baseline imbalance between treatment groups related to outcome (prior ototoxic treatment, age, sex, prior hearing loss or a combination of these): no prior ototoxic treatment, age and sex were well balanced, prior hearing loss unclear (baseline tests were performed, but results nm). *Difference in ototoxic drugs other than platinum analogue between treatment groups (furosemide, gentamycin, anthracyclines, vincristine)*: cumulative anthracycline dose nm, furosemide and gentamycin nm; vincristine not given. *Difference in cumulative platinum dose between treatment groups*: cumulative dose unclear, but according to protocol participants in both treatment groups should have received the same dose. *Difference in length of follow-up between treatment groups*: unclear (length of follow-up nm for either group). *Difference in impaired renal function at time of platinum treatment between treatment groups*: unclear. *An insensitive instrument was used to evaluate ototoxicity*: no (used pure tone audiometry).

Nm: not mentioned

GRADE assessments

The reasons for downgrading or upgrading should be included in the Summary of Findings tables. The level of evidence of each outcome/result should be included at the description of the evidence.

Discussion

Applicability of the evidence to clinical practice

Things to consider are:

- The quality of the evidence
- "No evidence of effect" is not the same as "evidence of no effect", but can be caused by for example a low power or a too short follow-up duration for the result to occur.
- Impact of missing data on the findings of the review
- Was a suitable test used to assess an outcome?
- When were included studies performed? Have there since been changes in treatment protocols including supportive care?
- External validity
- Keep in mind the clinically relevant effect (not only if a result has a P-value < 0.05)

More information:

- ✓ *Kremer LCM, van Dalen EC. P-values and confidence Intervals Evid.-Based Child Health 2008; 3(3): 904-906. Erratum in: Evid.-Based Child Health 2008;3(4): 1156*
- ✓ *Van Dalen EC, van der Pal HJ, Bakker PJ, Caron HN, Kremer LC. Cumulative incidence and risk factors of mitoxantrone-induced cardiotoxicity in children: a systematic review. Eur J Cancer 2004;40(5):643-652*

Strengths and limitations of the review

Here you can discuss for example if you used a broad search strategy. If that is the case, the presence of reporting bias is unlikely. If language restriction was used the risk of language bias should be considered.

Conclusions

Make sure that your conclusions are only based on the results of your review.

Implications for future research

Explain for example what study design is necessary, what types of patients should be included, that the study should have enough power and how outcomes should be defined.

3 Practical information regarding the organisation of the systematic review development

3.1. Authorship and roles

Authorship criteria and author order should be communicated at the start of the review process.

The members of the review team will be co-authors of the manuscript if they are substantially involved in the review process (i.e. both the review and the manuscript development). If not, their contribution should be included in the acknowledgements section of the manuscript. In the authorship guidelines of the International Committee of Medical Editors you can find more information about author requirements (http://www.icmje.org/).

In general, the first author coordinates the review development. He or she is one of the independent review authors performing study selection, data extraction, risk of bias assessment and GRADE assessment of all included studies and he or she performs the basic analyses. He or she drafts the first version of both the protocol and the manuscript of the completed review.

The systematic review & guideline unit will provide general methodological support and guide the review process. In turn, the involved members of the unit will become co-authors. More specialised help, like statistical or search related, needs to be obtained by other members of the review team.

3.2. Timeline for systematic review development

It is difficult to provide a timeline for systematic review development as it depends on many factors, like number of possible studies identified while running the search, number of eventually included studies and the time the review author team can dedicate to the review. Also, some journals require the protocol to be peer reviewed and published before starting the actual review. For example protocols for Cochrane systematic reviews are peer reviewed and subsequently published in the Cochrane Library. As a rough estimate you should count on 6 months before the protocol is finalized and 1 year to perform the review. Be aware that journals sometimes require the search not be older than a specific time period. This means that it might be necessary to perform a search update before submitting.

3.3 Updating the review

New studies can change the conclusions of the review and can thus have an impact on clinical practice. Although most journals do not require you to regularly update your systematic review (like every 2 or 5 years) it is something that should definitely be considered. For Cochrane systematic reviews it is always mandatory.

When performing an update, always check the methodology of the earlier review and, as better methods might have become available, adjust where necessary. This can mean that you need to go back to already included studies, for example when new risk of bias or GRADE criteria have become available.

Also check the existing search terms as databases might have slightly changed their terminology or new terms might have become available.

References

1. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from https://handbook-5-1.cochrane.org/
2. Leclercq E, Leeflang MMG, van Dalen EC, Kremer LCM. Validation of Search Filters for Identifying Pediatric Studies in PubMed. J Pediatr 2013;162(3):629-634
3. Van Dalen EC, Kremer LC. Inclusion of studies into a Cochrane review. Evid.-Based Child Health 2006; 1(4): 1349-1351
4. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 1998;17(24): 2815-2834. Erratum in: Stat Med 2004;23(11): 1817
5. Kremer LCM, van Dalen EC, Vandermeer B, Offringa M. Meta-analysis and heterogeneity. Evid.-Based Child Health 2010;5(1): 12-16
6. Rosmarakis ES, Soteriades ES, Vergidis PI, Kasiakou SK, Falagas ME. From conference abstract to full paper: differences between data presented in conferences and journals. FASEB J 2005 May;19(7): 673-680
7. Van Dalen EC, Kremer LCM. Subgroup analyses. Evid.-Based Child Health 2009; 4(2):1140-1141
8. Hanley JA, Lippman-Hand A. If nothing goes wrong, is everything all right? Interpreting zero numerators. JAMA 1983; 249(13): 1743-1745
9. GRADEprofiler 3.6.1. Hamilton (ON): McMaster University (developed by Evidence Prime), 2011
10. Van As JW, van den Berg H, van Dalen EC. Medical interventions for the prevention of platinum-induced hearing loss in children with cancer. Cochrane Database of Systematic Reviews 2019, Issue 5. Art. No.: CD009219. DOI: 10.1002/14651858.CD009219.pub5
11. Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. Lancet 2002;359(9303):341-345
12. Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. Evidence-Based Medicine Working Group. JAMA 1994;272(3):234-237
13. Mulder RL, Font-Gonzalez A, Hudson MM, et al. Fertility preservation for female patients with childhood, adolescent, and young adult cancer: recommendations from the PanCareLIFE Consortium and the International Late Effects of Childhood Cancer Guideline Harmonization Group. Lancet Oncol 2021; 22(2): e45-e56

Appendices:

Appendix 1 Example of a search strategy for medical interventions for the prevention of platinum-induced hearing loss in children with cancer (adjusted from [10]; PubMed version).

1. For Hearing loss, we used the following MeSH headings and text words in the original version of the review and the first update: Deafness OR hearing loss OR Loss, Hearing OR hearing disorder OR hearing disorders OR auditory OR hearing impairment OR hearing impairments OR hearing impairment* OR hear* OR audiology OR audiologic OR audiometry OR audiometr* OR audiogram OR audiography OR ototoxicology OR ototoxic* OR hypoacusis OR hypoacuses OR hypoacus* OR ototoxicity OR deaf* OR cochleotoxicity For the second and third update, we optimized this search strategy by excluding "hear*".
2. For Cisplatin, we used the following MeSH headings and text words: Cisplatin OR cis-Diamminedichloroplatinum(II) OR Platinum Diamminodichloride OR Diamminodichloride, Platinum OR cis-Platinum OR cis Platinum OR Dichlorodiammineplatinum OR cis-Diamminedichloroplatinum OR cis Diamminedichloroplatinum OR cis-Dichlorodiammineplatinum(II) OR Platinol OR Platidiam OR Platino OR NSC-119875 OR Biocisplatinum OR CDDP OR CACP OR cisplatin* OR abiplatin OR neoplatin OR cis-DDP
3. For Carboplatin, we used the following MeSH headings and text words: Carboplatin OR cis-Diammine(cyclobutanedicarboxylato)platinum II OR CBDCA OR Carbosin OR Pharmachemie Brand of Carboplatin OR Carbotec OR Columbia Brand of Carboplatin OR Ercar OR Almirall Brand of Carboplatin OR JM-8 OR JM 8 OR JM8 OR Neocarbo OR Neocorp Brand of Carboplatin OR NSC-241240 OR NSC 241240 OR NSC241240 OR Paraplatin OR Carboplat OR Paraplatine OR Bristol-Myers Squibb Brand of Carboplatin OR Platinwas OR Chiesi Brand of Carboplatin OR Ribocarbo OR ribosepharm Brand of Carboplatin OR Blastocarb OR Lemery Brand of Carboplatin OR Nealorin OR Prasfarma Brand of Carboplatin OR carboplatin*
4. For Oxaliplatin and other platinum compounds, we used the following MeSH headings and text words: Oxaliplatin OR oxaliplatin* OR 1,2-diamminocyclohexane(trans-1)oxolatoplatinum(II) OR oxaliplatine OR platinum(II)-1,2-cyclohexanediamine oxalate OR 1,2-diaminocyclohexane platinum oxalate OR oxalato-(1,2-cyclohexanediamine)platinum II OR cis-oxalato-(trans-l)-1,2-diaminocyclohexane-platinum(II) OR Eloxatine OR Eloxatin OR oxaliplatin, (SP-4-2-(1S-trans))-isomer OR oxaliplatin, (SP-4-3-(cis))-isomer OR ACT 078 OR ACT-078 OR oxaliplatin, (SP-4-2-(1R-trans))-isomer OR 63121-00-6 OR 61825-94-3 OR dacotin OR dacplat OR jm-83 OR l-ohp OR oxalatoplatinum OR rp 54780 OR sr-96669 OR Platinum OR Platinum Compounds OR platinum* OR organoplatinum compounds [mh]
5. For Children, we used the following MeSH headings and text words in the original version of the review and the first update: Infant OR infan* OR newborn OR newborn* OR new-born* OR baby OR baby* OR babies OR neonat* OR child OR child* OR schoolchild* OR schoolchild OR school child OR school child* OR kid OR kids OR toddler* OR adolescent OR adoles* OR teen* OR boy* OR girl* OR minors OR minors* OR underag* OR under ag* OR juvenil* OR youth* OR kindergar* OR puberty OR puber* OR pubescen* OR prepubescen* OR prepuberty* OR pediatrics OR pediatric* OR paediatric* OR peadiatric* OR schools OR nursery school* OR preschool* OR pre school* OR primary school* OR secondary school* OR elementary school* OR elementary school OR high school* OR highschool* OR school age OR schoolage OR school age* OR schoolage* OR infancy OR schools, nursery OR infant, newborn For the second and third update, we used the following MeSH headings and text words: infan* OR newborn* OR new-born* OR perinat* OR neonat* OR baby OR baby* OR babies OR toddler* OR minors OR minors* OR boy OR boys OR boyfriend OR boyhood OR girl* OR kid OR kids OR child OR child* OR children* OR schoolchild* OR schoolchild OR school child[tiab] OR school child*[tiab] OR adolescen* OR juvenil* OR youth* OR teen* OR under*age* OR pubescen* OR pediatrics[mh] OR pediatric* OR paediatric* OR peadiatric* OR school[tiab] OR school*[tiab] OR prematur* OR preterm* (Leclercq 2013 [2])

6. For RCTs/CCTs, we used the following MeSH headings and text words in the original version of the review and the first update: (Randomized controlled trial[pt] OR controlled clinical trial[pt] OR randomized[tiab] OR placebo[tiab] OR drug therapy[sh] OR randomly[tiab] OR trial[tiab] OR groups[tiab]) AND humans[mh] For the second and third update, we used the following MeSH headings and text words:(Randomized controlled trial[pt] OR controlled clinical trial[pt] OR randomized[tiab] OR placebo[tiab] OR drug therapy[sh] OR randomly[tiab] OR trial[tiab] OR groups[tiab]) NOT (animals [mh] NOT humans [mh]) (Higgins 2011 [1])
Final search 1 AND (2 OR 3 OR 4) AND 5 AND 6
*[pt = publication type; tiab = title, abstract; sh = subject heading; mh = MeSH term; *=zero or more characters; RCT = randomized controlled trial; CCT = controlled clinical trial]*

Appendix 2.

A. Example of an empty evidence table for observational studies

Clinical question				
First author et al. Title. Journal year;volume:pages				
Study design **Treatment era** **Follow-up**	**Participants**	**Treatment**	**Main outcomes**	**Additional remarks** **Risk of bias assessment**
Study design: Treatment era: Follow-up:	Type and number of participants: Diagnosis: Age at diagnosis: Age at testing/follow-up: Gender: Controls:	Chemotherapy: Irradiation: Surgery: Other treatments:	Outcome definitions: Results:	Additional remarks: Risk of bias: Selection bias: Attrition bias: Detection bias: Confounding:

B. Example of a filled in evidence table

What is the frequency of CAD and what are risk factors?				
Van der Pal HJ et al. High risk of symptomatic cardiac events in childhood cancer survivors. J Clin Oncol 2012; 30(13): 1429-37.				
Study design **Treatment era** **Follow-up**	**Participants**	**Treatment**	**Diagnostic test** **Main outcomes**	**Additional remarks**
Study design: Retrospective single-center cohort Treatment era: Diagnosed between 1-1-1966 and 1-1-1996 Follow-up: Median 22.5 years (elsewhere in the manuscript 22.2 years), range 5 to 44.5 years since primary cancer diagnosis	N=1362 5-year childhood cancer survivors Diagnosis: ALL N=302 (22.2%) ANLL N=30 (2.2%) Non-Hodgkin's disease N=167 (12.3%) Hodgkin's disease N=104 (7.6%) Nephroblastoma N=186 (13.7%) Soft tissue sarcoma N=131 (9.6%) Ewing sarcoma N=53 (3.9%) Osteosarcoma N=73 (5.4%) CNS tumor N=124 (9.1%) Neuroblastoma N=85 (6.2%) Germ cell tumor N=45 (3.3%) Other N=62 (4.5%) Age at diagnosis: 0-4 years: N=596 (43.7%) 5-9 years: N=378 (27.8%) 10-14 years: N=309 (22.7%) 15-18 years: N=79 (5.8%)	Chemotherapy: N=1167 (85.7%), of which anthracyclines N=565 (48.4%); cumulative dose range 1->500 mg/m^2; dose unknown in N=15. Other agents and doses not reported for all survivors. Irradiation: N=597 (43.8%), of which cardiac irradiation N=266 (44.6%); unknown in 1 patient. Cardiac irradiation was defined as: • Thorax (=left lung, mantle field, and/or mediastinum) N=84, dose in EQD2 median 24.08, range 9.47-88.46 • Abdomen (=whole abdomen, left kidney, inverted Y field, and/or PAO) N=65, dose in EQD2 median 26.9, range 3.73-57.19	Diagnostic test used for CAD assessment: Childhood Cancer Registry, medical records or general practitioners or attending physicians; all cardiac events were diagnosed by cardiologists and validated by a cardiologist. Timing of the diagnostic test: Time at risk started 5 years from diagnosis. Survivors who developed a cardiac event in the first 5 years after primary cancer diagnosis were eligible only if they had recovered (i.e. no symptoms of cardiac events or treatment) within the same 5 years. Survivors who did not recover within 5 years were excluded. Outcome definitions: Cardiac ischemia/infarction grade 3 or higher (i.e. symptomatic) according to the CTCAEv3 diagnosed more than 5 years after primary cancer diagnosis	Risk of bias: Selection bias: Low risk (study group consists of all patients included in the original cohort) Attrition bias: Low risk (complete follow-up) Detection bias: Unclear risk (no information on blinding of outcome assessors provided) Confounding: High risk (only univariable analyses available) Funding of the trial: Foundation of Pediatric Cancer Research Amsterdam

	Age at testing/follow-up: Median attained age 29.1 years, range 5.2 to 54.2 years Gender: 745 males (54.7%); 617 females (45.3%) Cardiovascular risk factors (like dyslipidemia, hypertension, obesity, inactivity, diabetes mellitus, smoking, genetic factors): Not reported Controls: No	• Spine N=89, dose in EQD2 median 30.14, range 8-50.11 • TBI N=28, dose in EQD2 median 15.75, range 14-21.60 Dose unknown in N=10 Chemotherapy only: N=658 (48.3%) Irradiation only: N=88 (6.5%) Chemotherapy and irradiation: N=509 (37.4%) Stem cell transplant: Not reported *N=107 no chemotherapy and/or radiotherapy (surgery only) (7.9%)* N=723 cardiotoxic therapy (=anthracyclines and/or cardiac radiotherapy) (53.1%)	Occurrence of CAD: N=6 (0.4%) Competing risk cumulative incidence with death from any cause or another cardiac event as competing risks: 40-year cause-specific cumulative incidence: 1.9% (95% CI 0 to 4.1%) Risk factors assessed: Yes Results of multivariable analyses: Not applicable Results of univariable analyses: Competing risk cumulative incidence with death from any cause or another cardiac event as competing risks: 40-years: Cardiotoxic therapy no: 0% Cardiotoxic therapy yes: 4.9% (95% CI 0 to 11.2%) Radiotherapy (=cardiac irradiation and no anthracyclines with or without all other treatment) no: 0.3% (95% CI 0 to 0.8%) Radiotherapy yes: 6% (0 to 13.3%)	

Appendix 3. Cochrane Childhood Cancer Risk of bias assessment criteria for observational studies*

	Internal validity
Study group	**Selection Bias (is the study group representative?)**: Low risk if: The study group consisted of more than 90% of the original cohort of eligible patients for the review for short-term outcomes and more than 75% for long-term outcomes *or* It was a random sample with respect to cancer treatment and important prognostic factors (i.e. …..).
Follow-up	**Attrition bias (is the follow-up adequate?)**: Low risk if: The outcome was assessed for more than 90% of the study group for short-term outcomes and more than 75% for long-term outcomes
Outcome	**Detection bias (are the outcome assessors blinded for important determinants related to the outcome?)**: Low risk if: The outcome assessors were blinded for important determinants related to the outcome
Risk estimation**	**Confounding (are the analyses adjusted for important confounders?)**: Low risk if: Important prognostic factors (i.e. ….) were taken adequately into account.

Each bias item should be scored as low risk, high risk or unclear risk (no overall scores should be calculated); attrition bias and detection bias should be scored for each outcome separately

***** based on previously described checklists according to evidence-based medicine criteria [11, 12]. For the Risk of bias assessment of case-control studies the criteria need to be slightly adapted with regard to the selection of cases and controls, which should be based on comparable patient characteristics.*

*** only applicable when risk factor analyses have been performed.*

Appendix 4. Example of a Cochrane Summary of Findings table including GRADE assessment for randomized studies [10]

Amifostine compared to no otoprotective intervention for the prevention of platinum-induced hearing loss in children with cancer						
Patient or population: children with cancer treated with platinum-based therapy **Settings:** paediatric oncology departments **Intervention:** amifostine **Comparison:** no otoprotective intervention						
Outcomes	**Illustrative comparative risks* (95% CI)**		**Relative effect (95% CI)**	**No of participants (studies)**	**Certainty of the evidence (GRADE)**	**Comments**
	Assumed risk	**Corresponding risk**				
	No otoprotective intervention	**Amifostine**				
Ototoxicity* (i.e. *hearing loss or tinnitus, or both)						
Ototoxicity according to NCICTCv2 criteria with intra-arterial platinum (combined asymptomatic and symptomatic disease) Exact test method not reported Follow-up not mentioned	**769 per 1000**[a]	**992 per 1000** (723 to 1000)	**RR 1.29** (0.94 to 1.77)	28 (1 study)	⨁⨁⊖⊖ **Low**[b,c]	When only looking at symptomatic disease there was also no significant difference between treatment groups (RR 0.87, 95% CI 0.14 to 5.32; GRADE assessment identical to combined asymptomatic and symptomatic disease analysis).
Ototoxicity according to NCICTCv2 criteria with intravenous platinum (combined asymptomatic and symptomatic disease) Objective and subjective audiometric evaluations were	**789 per 1000**[a]	**821 per 1000** (600 to 1000)	**RR 1.04** (0.76 to 1.44)	36 (1 study[d])	⨁⨁⊖⊖ **Low**[c,e]	For 3/39 children included in the study (all in the amifostine group) there were no data on ototoxicity. The RR reported here resulted from the available-data analysis. Intention-to-treat analyses (i.e. best-case and worst-case scenarios) also showed no significant difference between the treatment groups. When only looking at symptomatic disease there was also no significant difference between treatment

performed, no further information provided Follow-up not mentioned						groups (available-data analysis: RR 0.87, 95% CI 0.14 to 5.32; intention-to-treat analyses (i.e. best-case and worst-case scenarios) also showed no significant difference between treatment groups). The GRADE assessment for the worst-case and best-case scenarios and the symptomatic disease-only analysis was identical to that of the 'available-data' analysis for the combined asymptomatic and symptomatic disease analysis.
Ototoxicity according to modified Brock criteria with intravenous platinum (combined asymptomatic and symptomatic disease) Audiograms were performed, but no further information provided Follow-up not mentioned	**556 per 1000**[a]	**594 per 1000** (411 to 861)	**RR 1.07** (0.74 to 1.55)	82 (1 study)	⨁⨁⊖⊖ **Low**[c,f]	It should be noted that these 82 children were part of a larger study group; they were considered in a special interim analysis of the incidence of toxicity. The total number of eligible participants was unclear and as a result we were unable to perform an intention-to-treat analysis. Also, we were unable to check if the ototoxicity results were available for at least 50% of the eligible participants. In the 'Methods' section, we stated that if that was not the case, we would not report the results due to the associated high risk of attrition bias. However, we decided to give this study the benefit of the doubt. When only looking at symptomatic disease, there was also no significant difference between treatment groups (RR 1.00, 95% CI 0.57 to 1.75; GRADE assessment identical to combined asymptomatic and symptomatic disease analysis).
Survival						
Survival (overall and event-free) – not reported	—	—	—	—	—	No information on overall and event-free survival
Tumour response						

Tumour response with intra-arterial platinum (good and partial remission) Follow-up not mentioned	**583 per 1000**[a]	**933 per 1000** (566 to 1000)	**RR 1.6** (0.97 to 2.63)	27 (1 study)	⨁⨁⊖⊖ **Low**[b,c]	For 1/28 children included in the study (in the control group) there were no data on tumour response. The RR reported here resulted from the available-data analysis. Intention-to-treat analyses also showed no significant difference between the treatment groups in the best-case scenario, but in the worst-case scenario there was a significant difference in favour of amifostine (GRADE assessment identical to available-data analysis). Due to the nature of this outcome (number of participants with a remission) a high event rate is favourable. The studies using intravenous platinum did not report on this outcome.
Adverse effects other than ototoxicity						
Renal toxicity/vomiting/ cardiotoxicity (all grade ≥ 3 according to NCICTCv2 criteria) with intra-arterial platinum Follow-up not mentioned	Renal toxicity: no significant difference between treatment groups (Fischer's exact test P = 0.21) Vomiting: significant difference in favour of the control group (RR 9.04, 95% CI 1.99 to 41.12) Cardiotoxicity: none of the participants in this study experienced cardiac toxicity grade 3 or 4)		—	28 (1 study)	⨁⨁⊖⊖ **Low**[b,c]	The studies using intravenous platinum did not provide adequate data on adverse effects.
Quality of life						
Quality of life – not reported	—	—	—	—	—	No information on quality of life
*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; **NCICTCv2:** National Cancer Institute Common Toxicity Criteria version 2; **RR:** risk ratio.
GRADE Working Group grades of evidence **High certainty:** further research is very unlikely to change our confidence in the estimate of effect. **Moderate certainty:** further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate. **Low certainty:** 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 certainty:** we are very uncertain about the estimate.
[a]The assumed risk was based on the prevalence in the control group of the included study. [b]Presence of selection bias, performance bias, detection bias and other bias was unclear; low risk of attrition bias and reporting bias (downgraded one level). [c]As this was a small study with a total number of events fewer than 300 (the threshold rule-of-thumb value stated in the GRADEpro software (GRADEpro)), we downgraded one level. [d]This was a controlled clinical trial. [e]High risk of selection bias, performance bias, attrition bias and reporting bias; unclear risk of detection bias and other bias (downgraded one level). [f]High risk of attrition and reporting bias; unclear risk of selection bias, performance bias and other bias; low risk of detection bias (downgraded one level).

Appendix 5. Example of a Summary of Findings table including GRADE assessment for observational studies [13]

Outcome	Study	No. of participants	Follow up (median/mean, range) yr	Alkylating agents Radiotherapy to ovaries	Events	Effect size	Risk of bias
1.5 Risk POI after procarbazine (n=4 studies)	Chemaitilly 2006*	3,390 CCS	>5 yr after cancer diagnosis	Alkylating agents: 49.7%; Radiotherapy to ovaries: 24.5%	215/3390 (6.3%) amenorrhea within 5 yr after their cancer diagnosis	*Odds ratio (95% CI) for amenorrhea age at diagnosis 0-12 yr* Procarbazine yes vs. no: OR 3.2 (1.3-7.3) *Odds ratio (95% CI) for amenorrhea age at diagnosis 13-20 yr* Procarbazine yes vs. no: OR 2.6 (1.4-4.7)	SB: high risk AB: low risk DB: unclear CF: low risk
	Thomas-Teinturier 2013*	706 CCS	>5 yr after cancer diagnosis	Any alkylating agent: 47.7%; Procarbazine: 7.2%; Radiotherapy to ovaries: 56.7%	62/706 (8.9%) nonsurgical menopause; 15/706 (2.1%) nonsurgical premature menopause <age 40 yr	*Relative risk (95% CI) for nonsurgical menopause* Procarbazine dose per g/m^2: RR 2.5 (1.4-5.8)	SB: high risk AB: low risk DB: unclear CF: low risk
	Thomas-Teinturier 2015*	108 CCS	>3 yr after cancer treatment	Any alkylating agent: 100%; Procarbazine: 21.9%; Radiotherapy to ovaries: 17.6%	8/108 (7.6%) altered ovarian function (↑ FSH, ↓ AMH and amenorrhoea)	*Mean FSH* Procarbazine dose: β 0.012, p<0.001; (Each unit increase in procarbazine dose, mean FSH values increased by 0.012 IU/L)	SB: high risk AB: low risk DB: unclear CF: low risk
	Levine 2018*	2,930 CCS	>5 yr after cancer diagnosis	Alkylating agents: 46.5%; Procarbazine: 201 (7.2%);	110/2930 (3.8%) nonsurgical premature menopause <age 40 yr	*Odds ratio (95% CI) for nonsurgical premature menopause*	SB: high risk AB: low risk DB: unclear CF: low risk

			Radiotherapy to ovaries: 55.4%	Procarbazine dose <4000 mg/m^2 vs. 0: OR 3.07 (0.76-12.43); Procarbazine dose ≥4000 mg/m^2 vs. 0: OR 8.96 (5.02-16.00)
Quality of evidence:				
Study design:	+4	Retrospective cohort studies		
Study limitations:	-1	Limitations: Selection bias high in 4/4; Attrition bias low in 4/4; Detection bias unclear in 4/4; Confounding low in 4/4		
Consistency:	0	No important inconsistency, all show effect of procarbazine		
Directness:	0	Results are direct, population and outcomes broadly generalizable		
Precision:	0	No important imprecision, large sample size, high total number of events and narrow confidence intervals		
Publication bias:	0	Unlikely		
Effect size:	0	No large magnitude of effect		
Dose-response:	+1	Dose response relationship as higher doses are associated with an increased risk as compared to lower doses		
Plausible confounding:	0	No plausible confounding		
Quality of evidence:		⊕⊕⊕⊕ HIGH		
Conclusion:		Increased risk of POI after procarbazine vs. no procarbazine in female cancer survivors diagnosed before age 25 years. (4 studies significant effect; 7,134 participants; 395 events; 4 multivariable analyses)		

Abbreviations: AB, attrition bias; AMH, anti-Müllerian hormone; CCS, childhood cancer survivors; CF, confounding; DB, detection bias; FSH, follicle stimulating hormone; SB, selection bias; yr, year.

** Overlap in included patients in studies of Chemaitily 200 and Levine 2018; and Thomas-Teinturier 2013 and 2015.*