

BMJ Open Development of an algorithm to provide awareness in choosing study designs for inclusion in systematic reviews of healthcare interventions: a method study

Frank Peinemann,^{1,2} Jos Kleijnen^{1,3}

To cite: Peinemann F, Kleijnen J. Development of an algorithm to provide awareness in choosing study designs for inclusion in systematic reviews of healthcare interventions: a method study. *BMJ Open* 2015;**5**:e007540. doi:10.1136/bmjopen-2014-007540

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2014-007540>).

Received 24 December 2014
Revised 25 May 2015
Accepted 16 June 2015



¹Maastricht University, School for Public Health and Primary Care, Maastricht, The Netherlands

²Children's Hospital, University of Cologne, Cologne, Germany

³Kleijnen Systematic Reviews Ltd, York, UK

Correspondence to
Frank Peinemann;
pubmedprjournal@gmail.com

ABSTRACT

Objectives: To develop an algorithm that aims to provide guidance and awareness for choosing multiple study designs in systematic reviews of healthcare interventions.

Design: Method study: (1) To summarise the literature base on the topic. (2) To apply the integration of various study types in systematic reviews. (3) To devise decision points and outline a pragmatic decision tree. (4) To check the plausibility of the algorithm by backtracking its pathways in four systematic reviews.

Results: (1) The results of our systematic review of the published literature have already been published. (2) We recaptured the experience from our four previously conducted systematic reviews that required the integration of various study types. (3) We chose length of follow-up (long, short), frequency of events (rare, frequent) and types of outcome as decision points (death, disease, discomfort, disability, dissatisfaction) and aligned the study design labels according to the Cochrane Handbook. We also considered practical or ethical concerns, and the problem of unavailable high-quality evidence. While applying the algorithm, disease-specific circumstances and aims of interventions should be considered. (4) We confirmed the plausibility of the pathways of the algorithm.

Conclusions: We propose that the algorithm can assist to bring seminal features of a systematic review with multiple study designs to the attention of anyone who is planning to conduct a systematic review. It aims to increase awareness and we think that it may reduce the time burden on review authors and may contribute to the production of a higher quality review.

INTRODUCTION

When evaluating healthcare interventions, different categories of intervention such as medicinal versus non-medicinal therapy and different categories of outcomes such as intended effects, adverse events or

Strengths and limitations of this study

- We developed an algorithm to provide guidance for allocating various study designs to specific research questions. This can be viewed as a response to the lack of comprehensive guidance in major published methods documents.
- The terms used for defining the critical decision points of the algorithm, such as length of follow-up, frequency of events and types of outcome need to be interpreted in the context of the disease.
- Disease-specific circumstances and aims of interventions always have to be taken into account during application of the algorithm.
- We could follow and confirm the appropriateness of the pathways during the application of the algorithm on four selected systematic reviews.
- The checking of the plausibility of the algorithm was based on systematic reviews that were already completed. This approach is far from the everyday working condition and the approach may be biased by subjective expectations of the authors. We encourage independent evaluation of the algorithm.

health-related quality of life may sometimes be best answered by multiple study designs. Some designs have features, which preferably match the requirements of specific parts of a research question. Exclusively using data from randomised trials (RCTs) to evaluate whether an intervention might work has a number of limitations.¹ For example, RCTs may not be appropriate to estimate the incidence of rare (adverse) events. Other study designs, for example registry analyses, may incorporate the data of much more participants. These analyses may therefore complement the information on rare but important events. We have gathered some examples of research questions that cannot or only with

great difficulty be investigated in RCTs (table 1). A practical concern may arise with low numbers of patients with a rare disease. It might be difficult to conduct an RCT to evaluate patients with acquired severe aplastic anaemia. An ethical concern may arise with the treatment of severe diseases or life-threatening treatments. It might be obsolete to conduct an RCT to evaluate a new experimental treatment of patients with pancreatic cancer.

The methods of conducting systematic reviews of healthcare interventions are major components of 'evidence-based' medicine (EBM). In 2000, Sackett *et al*² defined EBM as the integration of 'best' research evidence with individual clinical expertise, patient values and expectations, and 'best' external research evidence. This definition may be visualised by three overlapping circles in a Venn diagram.³ The area of intersection, where all three different resources meet, should represent the EBM. To classify more valid and less valid information, the 'levels of evidence' specify a hierarchical order for various research designs based on their internal validity. The highest level of valid data, that is, the 'best' evidence, however, is not always available. In table 2, we present the classification of some of the major study designs for intended effects of therapy. There appears to be some variation in the hierarchy among some authors and institutions issuing 'evidence-based' guidelines or systematic reviews. All authors agree that RCTs have the highest 'level of evidence' with respect to minimising the risk of bias. The prospective non-randomised controlled clinical trial (CCT) has an

experimental design and its internal validity should be regarded lower than a randomised trial but higher than an observational study. Prospective cohort studies have a potential for a lower risk of bias than retrospective cohort studies because they have lower risk of recall bias and confounding.¹¹ In cohort studies groups are defined by exposure whereas in case-control studies groups are defined by outcome status. Both points are acknowledged in the 'hierarchy of evidence' by some but not all authors. Case series and case reports are descriptions of one or more individual cases. Some authors combine both designs in one category while others place case series a higher level.

In figure 1 we show a study design classification tree including the main features of study designs that makes them distinct from others conforming with the reports of the Centre for Evidence-Based Medicine (CEBM), the National Institute for Health and Care Excellence (NICE) and the Centre for Reviews and Dissemination (CRD).^{7 12 13} Examples of distinguishing study characteristics are the concurrent versus the historical control group or the participants being or not being allocated to the treatment groups by the investigator. Within group comparison has also been referred to as a before and after study in a previous CRD report.

We observed a lack of a clear and comprehensive guidance to optimise the choice of study designs in systematic reviews. A simple and clear algorithm could remind authors instantly about important issues that should be considered. Consequently, we aimed to develop such an algorithm. The algorithm could raise awareness of the

Table 1 Research questions that cannot or that can only with difficulty be investigated in RCTs

Topic	Reason
Research questions that in certain circumstances cannot be investigated in RCTs	
Life-threatening intervention, for example, intervention with high early treatment-related mortality	Allocation to intervention group endangers life
Certain second-line interventions reserved for refractory patients that did not respond to first-line standard therapy	Ultimo ratio and therefore no control group by definition. Example: Haematopoietic stem cell transplantation from unrelated versus related donors for patients with acquired severe aplastic anaemia
Pregnant women	Ethical concerns against inclusion in experiments
Infants	Ethical concerns against inclusion in experiments
Interventions that have been shown to produce a dramatic effect	The magnitude of benefit of one particular intervention such as insulin to treat diabetes mellitus would render any intervention a neglect of healthcare if insulin would be omitted unless the new treatment does also have a dramatic effect
Lack of consent to participate	Cheating persons is not legal
Studies that do not comply with the Declaration of Helsinki	The set of ethical principles regarding human experimentation is regarded as the cornerstone document of human research ethics
Research questions that can with difficulty be investigated in RCTs	
Rare adverse events and other rare safety outcomes	Number of study participants is too low
Allocation of alternative interventions is dominated by patients' preferences	Treatment group is chosen by a patient because of specific expectations of effectiveness, adverse events or health-related quality of life

RCTs, randomised controlled trials.



Table 2 Definition, classification and hierarchy of study designs for intended effects of therapy

Category	Contr	Prosp	Design	Description	Evidence level by some institutions or authors							
					I	II	III	IV	V	VI	VII	VIII
Experimental	Yes	Yes	Randomised controlled trial	Random, concealed allocation of participants to an intervention and a control group	1	1	1	1	1	1	1	1
Experimental	Yes	Yes	Prospective non-randomised controlled clinical trial	The method of allocation by the researcher falls short of genuine randomisation and fails to conceal the allocation sequence	2	2	NR	NR	2	1	NR	NR
Observational	Yes	Yes	Prospective cohort study	Comparison of outcome rates between treatment groups (intervention vs comparator)	3	2	2	2	3	2	2	2
Observational	Yes	Yes	Nested case-control study	Case-control study nested in a prospective cohort study, combines advantage of two study designs	NR	NR	NR	NR	NR	NR	NR	NR
Observational	Yes	No	Retrospective cohort study	Comparison of outcome rates between treatment groups (intervention vs comparator)	3	3	3	2	3	2	2	2
Observational	Yes	No	Case-control study	Comparison of treatment rates between outcome groups (cases vs controls)	4	4	3	3	3	2	2	2
Observational	No	No	Registry analysis	Description of the outcome in many patients collected with a wide range of settings and patients' characteristics	NR	NR	NR	NR	NR	NR	NR	NR
Observational	No	No	Case series	Description of the outcome in a number of 1 or more cases of an intervention	5	5	4	4	3	3	3	3
Observational	No	No	Case report	Description of the outcome in a number of 1 or more cases of an intervention	5	5	5	NR	NR	4	3	3
Observational	No	No	Health outcomes research		NR	NR	NR	2	NR	NR	NR	NR
Observational	No	No	Ecological study		NR	NR	NR	2	NR	NR	NR	NR
Observational	No	No	Cross-sectional study	Intervention and outcome data collected at one particular time.	5	NR	NR	NR	NR	3	NR	NR
Observational	No	No	Within-group comparison	Also known as before-and-after study. Comparison of outcomes before and after an intervention.	5	NR	NR	NR	2	3	NR	NR
Others	NA	No	Expert opinion		5	NR	5	5	NR	4	4	4
Others	NA	No	Consensus recommendation		5	NR	5	NR	NR	NR	4	NR
Others	NA	No	Pathophysiological study		5	NR	5	5	NR	4	NR	NR
Others	NA	No	Animal study		5	NR	NR	NR	NR	NR	NR	NR

Experimental: In an experimental study, the researcher allocates participants to different treatment groups. Observational: In an observational study, the participants are not allocated by the researcher. Control: yes: with control group (comparative study), no: no control group (single-arm study). Evidence level by some institutions or authors: I: Present review; II: Vandembroucke 2008;⁴ III: Gemeinsamer Bundesausschuss (G-BA) Federal Joint Committee 2013;⁵ Centre for Evidence-based Medicine (CEBM) 2009;⁶ V: Centre for Reviews and Dissemination (CRD);⁷ VI: Khan 2011;⁸ VII: National Institute for Health and Clinical Excellence (NICE): levels of evidence were specified in 2004⁹ but not in the updated versions in 2008 and 2013; VIII: Scottish Intercollegiate Guidelines Network (SIGN) 2011.¹⁰ CCS, case-control study; Contr, control group; CR, case report; CS, case series; HOR, Health outcome research; NA, not applicable; NCC, nested case-control study; NRCCT, non-randomised controlled clinical trial; NR, not reported; PCS, prospective cohort study; Prosp, prospective design; RCS, retrospective cohort study; RCT, randomised controlled trial.

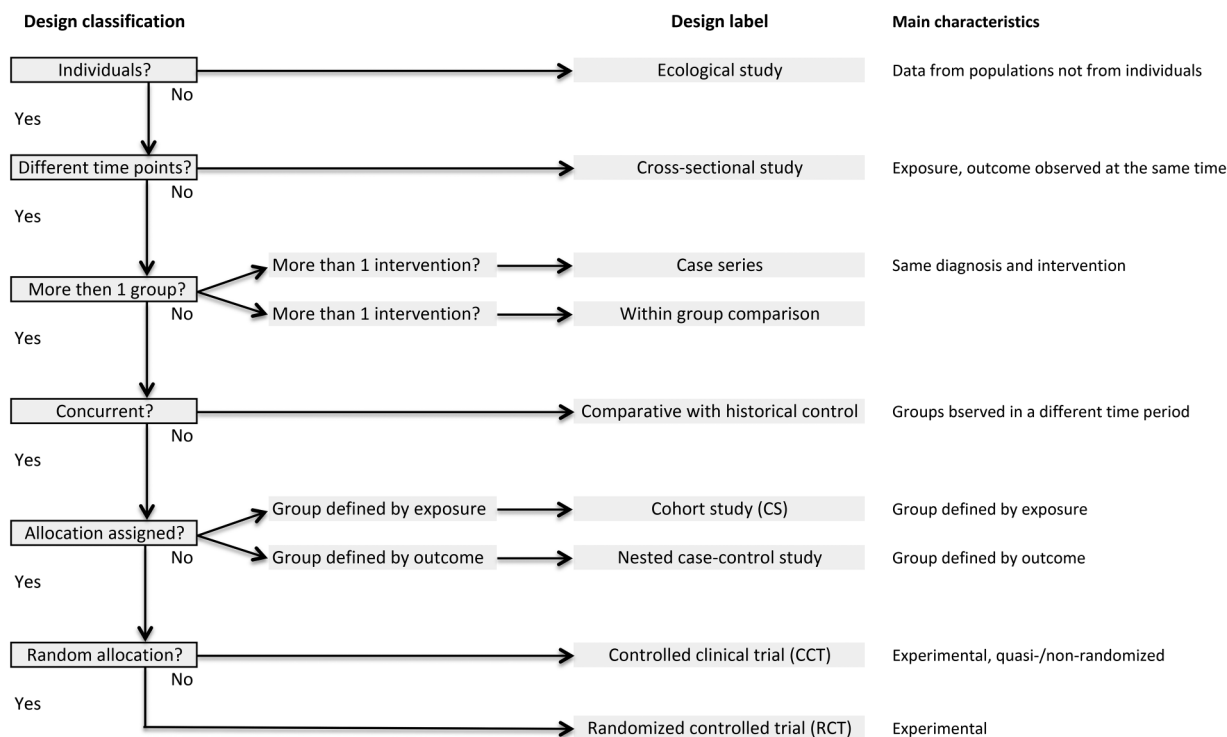


Figure 1 Decision algorithm to help define study designs.

prospects of multiple study designs, especially for less experienced authors. It could bring attention to issues that could be overlooked but are pertinent to healthcare. We expect that it may reduce the time burden for review authors and may facilitate the production of higher quality reviews.

METHODS

The objectives of the study may be subdivided into four items, which are used to structure the text.

Objective 1

First, we systematically reviewed the literature about the advantages and disadvantages of integrating multiple study designs in systematic reviews. Criteria for considering and search methods for identification of publications and data collection and analysis are described in a precursor article that is firmly connected to the present paper.¹⁴ These results form the information base for the topics of the current paper.

Objective 2

Second, we have conducted systematic reviews that are associated with the integration of multiple study designs. We reflected what could be learnt from this experience and what could also be a helpful piece of information for upcoming authors. Experience gathered in these papers was weaved into checking the plausibility of the algorithm.

Objective 3

Third, we wanted to know if major not-for-profit publishers of systematic reviews have included guidelines on the integration of multiple study designs in their manuals. We non-systematically searched the internet sites of 12 selected high-profile institutions and we transferred the relevant statements. The access dates are provided with the reference. We conceived an idea, how a decision tree could look like that should contain major characteristics of clinical studies, provide easy to follow pathways and close on recommended study designs. The resulting algorithm should depict the necessary information in a clear and straightforward way and combine all parts on a single page. We observed that the length of follow-up and the frequency of events are essential components of every outcome assessment and we introduced those items as binary decision points. Furthermore, we were convinced on theoretical grounds that both components are critical in the process of choosing the appropriate study design. Fletcher¹⁵ classified outcomes into the following five simple categories: death, disease, discomfort, disability and dissatisfaction (table 3) and we judged that this classification fits well into the algorithm. We did not consider economic outcomes. We selected study design labels from the Cochrane Handbook to maintain a common language and a reference for its the descriptions.¹⁶ We combined similar design concepts of the Cochrane Handbook that appeared redundant for the purpose of the algorithm. For example, the experimental design comprises mainly two types of design, the

Table 3 Outcomes of disease

Outcome	Description	Type of outcome
Death	A bad outcome if untimely	Investigator reported
Disease	A set of symptoms, physical signs, and laboratory abnormalities	Investigator reported
Discomfort	Symptoms such as pain, nausea, dyspnoea, itching, and tinnitus	Participant-reported disease-related symptoms
Disability	Impaired ability to go about usual activities at home, work, or recreation	Participant-reported disease-related impaired function
Dissatisfaction	Emotional reaction to disease and its care, such as sadness or anger	Participant-reported disease-related bother about impaired function and generic health-related quality of life

randomised controlled trial (RCT) and the non-randomised controlled trial (CCT).¹⁷ We used the term ‘CCT’ to combine the quasi-randomised controlled trial, the non-randomised controlled trial and the controlled before-and-after study. We used the term ‘cohort studies’ to combine the prospective and the retrospective cohort study designs. We used the term ‘case series’ to combine the case series study design and the uncontrolled before-and-after comparison design. The term ‘registry analyses’ is not listed in the Cochrane Handbook, though it may be classified as a retrospective subtype of cohort studies. Registries generally collect data that are confined to a specific disease, a specific intervention or a specific outcome. One example are the registry-based studies of the European Society for Blood and Marrow Transplantation, which registers data from transplanted patients after having received bone marrow or haematopoietic stem cells.¹⁸ Therefore, we wanted to accentuate this type of data selection and analysis and introduced the term ‘registry analyses’ into the algorithm. We did not consider the historically controlled design because changes over time are expected to because serious systematic differences between treatment groups. We also did not consider the cross-sectional study design due to the lack of observation over time.

Objective 4

Fourth, we checked the plausibility of the algorithm by backtracking its pathways with four previously conducted systematic reviews. The first author of the present paper was also the leading author of these systematic reviews and could apply his knowledge about all the details of the history of these systematic reviews. The second author checked whether the results of the plausibility check appeared sensible. The structured research question of each systematic review and the simulated pathways for choosing multiple study designs were described in detail. For this purpose, we structured the information such as the inclusion criteria by using the PICOTS-SD typology: participants (P), interventions (I), comparators (C) and outcomes (O), timing (T), setting (S), and study design (SD).^{7 19–21}

RESULTS

Objective 1

We identified ‘49 studies that compared the effect sizes between randomised and non-randomised controlled trials, which were statistically different in 35%’.¹⁴ We concluded: ‘The risk of presenting uncertain results without knowing for sure the direction and magnitude of the effect holds true for both non-randomised and randomised controlled trials. The integration of multiple study designs in systematic reviews is required if patients should be informed on the many facets of patient relevant issues of healthcare interventions’.¹⁴

Objective 2

The following four systematic reviews were used. The PICOTS-SD frames of these papers are shown in table 4.

Example 1: Non-rhabdomyosarcoma soft tissue sarcomas.²² This systematic review evaluated autologous haematopoietic stem cell transplantation (autoHSCT) following high-dose chemotherapy (HDCT) versus standard-dose chemotherapy (SDCT) in patients with non-rhabdomyosarcoma soft tissue sarcomas (NRSTS).

Example 2: Acquired severe aplastic anaemia.²⁴ This systematic review evaluated allogeneic haematopoietic stem cell transplantation (alloHSCT) from matched sibling donors (MSD) versus immunosuppressive therapy (IST) in patients with acquired severe aplastic anaemia (SAA).

Example 3: Localised prostate cancer.²⁵ This systematic review evaluated permanent interstitial low-dose-rate brachytherapy (LDR-BT) versus radical prostatectomy (RP) versus external beam radiotherapy (EBRT) and no primary therapy (NPT) in patients with localised prostate cancer.

Example 4: Negative pressure wound therapy.²⁷ This systematic review evaluated negative pressure wound therapy (NPWT) versus standard wound dressing in patients with wounds.

Objective 3

We looked at a sample of 12 high profile not-for-profit publishers of systematic reviews detailed in table 5. Of these, 10 have published guidance about their

Table 4 PICOTS-SD frames of the included systematic reviews

Examples	References	P	I	C	O	T	S	SD
Example 1: Non-rhabdomyosarcoma soft tissue sarcomas (NRSTS)	22 23	Patients with high-risk NRSTS	High-dose chemotherapy (HDCT) followed by autologous haematopoietic stem cell transplantation (autoHSCT)	Standard-dose chemotherapy (SDCT)	Overall survival (OS), treatment-related mortality (TRM)	5-year follow-up (FU)	Units in university hospitals specialised in transplantation	Randomised controlled trials (RCT)
Example 2: Acquired severe aplastic anaemia (SAA)	24	Patients with acquired SAA	Allogeneic haematopoietic stem cell transplantation (alloHSCT) from HLA-matched related donors	Immunosuppressive therapy (IST) using ciclosporin A (CSA) and antithymocyte globulin (ATG)	OS, treatment-related mortality (TRM)	5-year FU	Units in university hospitals specialised in transplantation	RCT, comparative clinical studies
Example 3: Localised prostate cancer	25 26	Patients with localised prostate cancer	Permanent interstitial low-dose rate brachytherapy (LDR-BT)	Radical prostatectomy (RP), external beam radiotherapy (EBRT), or no primary therapy (NPT)	OS, function and bother as well as health-related quality of life	5-year FU	Surgery and radiotherapy units in general hospitals	RCT
Example 4: Negative pressure wound therapy (NPWT)	27	Patients with chronic wounds	NPWT	conventional gauze dressing	Complete wound closure, severe adverse events such as bleeding	6-month FU	General hospitals	RCT

Table 5 Methods guidance by publishers of systematic reviews

ID (Reference)	Country	Name of Institution	Title of handbook
AHRQ 2009 ²⁸	USA	Agency for Healthcare Research and Quality	Methods (section of a completed report)
ASERNIP-S 2009 ²⁹	Australia	Australian Safety and Efficacy Register of New Interventional Procedures—Surgical; Royal Australasian College of Surgeons	General Guidelines for Assessing, Approving and Introducing New Surgical Procedures into a Hospital or Health Service
CADTH 2003 ³⁰	Canada	Canadian Agency for Drugs and Technologies in Health	Guidelines for Authors of CADTH Health Technology Assessment Reports
CEBM 2014 ¹²	UK	Centre for Evidence-Based Medicine	Study designs
Cochrane 2011 ¹⁶	UK, World	The Cochrane Collaboration	Cochrane Handbook for Systematic Reviews of Interventions; V.5.1.0; (updated March 2011)
CRD 2011 ⁷	UK	Centre for Reviews and Dissemination	Systematic Reviews. CRD's guidance for undertaking reviews in healthcare
HAS 2007 ³¹	France	French National Authority for Health	General method for assessing health technologies
IQWiG 2013 ³²	Germany	Institute for Quality and Efficiency in Health Care	Methoden V.4.0
MRC 2008 ³³	UK	Medical Research Council	Developing and evaluating complex interventions: new guidance
MSAC	Australia	Medical Services Advisory Committee	No handbook found
NICE 2013 ¹³	UK	National Institute for Health and Clinical Excellence	Guide to the Methods of Technology Appraisal
OHTAC	Canada	Ontario Health Technology Advisory Committee	No handbook found

ID, Identifier.

methodological procedure for preparing the systematic reviews.^{7 12 13 16 28–33} A range of other books or guidance documents on systematic reviews exist.^{8 34–38} We extracted the major statements of their methods guidance with respect to choosing the appropriate research design in online supplementary table S1. We did not identify an algorithm or a comprehensive guidance focusing on finding the appropriate research design in any of these methods guidance documents. We propose an algorithm, which is shown in figure 2. The algorithm has four decision points.

First, it should be decided whether the outcomes are typically evaluated at an early or late time point after start of treatment. The cut-off between a short and long follow-up depends on the type of disease, intervention and outcome and we list some examples that range from 30 days to 5 years (table 6).

Second, it should be decided whether the events of interest are regarded as rare or frequent. The cut-off between a rare and frequent event depends on the type of disease, intervention and outcome and we list some examples in table 7. A rare disease may be defined according to the Office of Rare Diseases Research (ORDR): “In the United States, a rare disease is generally considered to be a disease that affects fewer than 200 000 people”.⁴⁹

Third, the type of outcome of interest needs to be considered. We list some examples of outcomes, which may depend on length of follow-up and frequency of events (table 8). Additional examples for outcomes of respiratory tract disease are shown in table 9.

Fourth, the recommended study design for inclusion in a systematic review is assigned. We used the following

study design labels: RCT, CCT, nested case–control study, cohort study, case–control study, case series, case report and registry analysis. These and alternative study design labels are described in table 2.

Practical or ethical concerns may emerge as reasons to over-ride the earlier decisions or to switch to a more appropriate study design. We remind the reader at the bottom of figure 2 to reconsider the chosen path. This part is introduced to facilitate a flexible handling of the algorithm. Examples are shown in table 8. Ethical concerns are primarily associated with objections against experimental allocation.

Objective 4

We conducted a plausibility check of the algorithm's pathways by backtracking four own previously published systematic reviews. We marked the pathways by boxes that are filled in with a coloured background or that have a coloured frame lines.

Example 1: Non-rhabdomyosarcoma soft tissue sarcoma. In the first version of this systematic review, we assumed a long follow-up and frequent events regarding death by disease or complication: NRSTS in figure 3.²³ Thus, an RCT would be the best choice for all outcomes but we did not find any RCT and we did not find any comparative study. Instead, we identified only single-arm studies. We estimated overall survival and described adverse events but were unable to draw conclusions on the benefit of the intervention of interest. In a planned update 2 years later, we were able to identify a single RCT.²² Using different study types provided the advantage to report estimates of overall survival in the first version when RCTs were lacking. The advantage affected

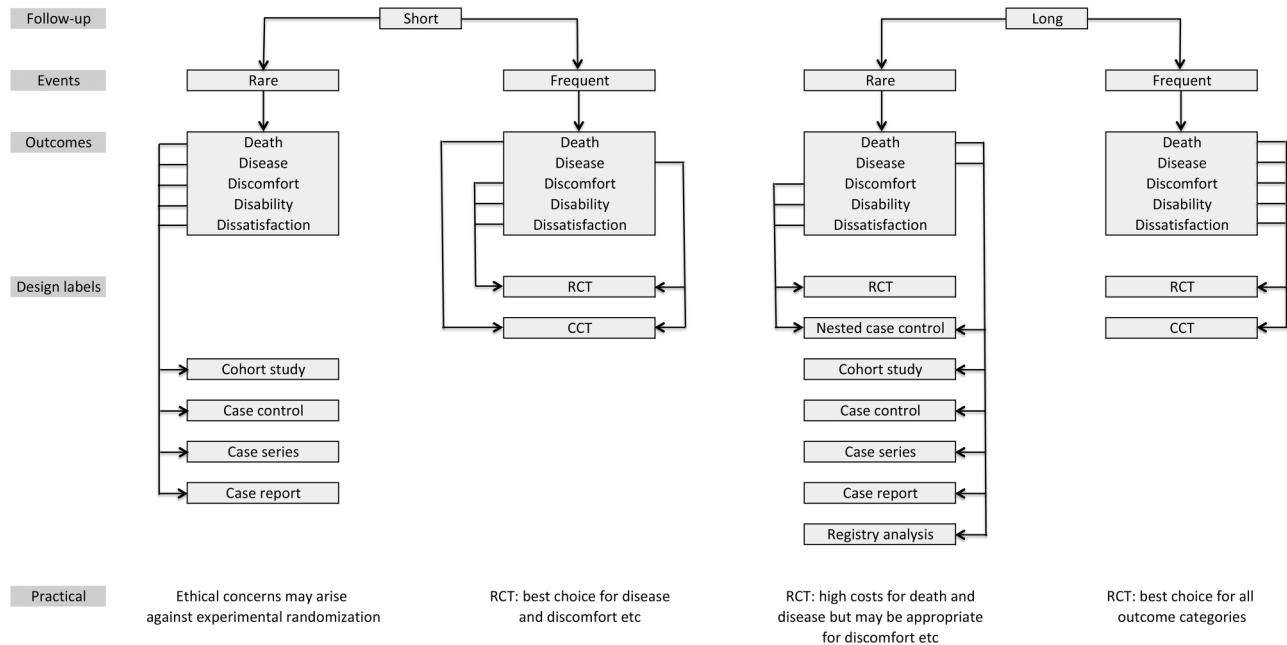


Figure 2 Algorithm Explanation:

- ▶ RCT: prospective randomised controlled trial, allocating experimental units via random assignment to a treatment or control condition with concealment of the allocation procedure
- ▶ CCT: prospective non-randomised controlled clinical trial, allocating experimental units via non-random assignment to a treatment or control condition; discomfort etc: discomfort, disability and dissatisfaction
- ▶ Nested case-control study: case-control study nested within a prospective, observational cohort study
- ▶ Cohort study: prospective, observational cohort study
- ▶ Case series: retrospective, observational tabulation of data from participants without consecutive enrolment
- ▶ Case report: retrospective, observational report of data from one participants up to three patients
- ▶ Registry analysis: retrospective, observational analysis of participants data from various sources transferred to and collected in a database CCT, controlled clinical trial; nested case contr: nested case-control study within a prospective cohort study; RCT, prospective randomised controlled trial, allocating experimental units via random assignment to a treatment or control condition with concealment of the allocation procedure

also the update version because the reporting of adverse events exceeded the scope of a single RCT considerably.

Example 2: Acquired severe aplastic anaemia. In this systematic review, we assumed a long follow-up and frequent events regarding death by disease or complication: SAA in [figure 3](#).²⁴ Thus, an RCT would be the best choice for all outcomes. As we did not identify any RCT, we included other comparative study designs. We tried to lower the risk of bias imposed by the non-randomised design. Eligible studies needed to be prospective non-randomised controlled trials, to meet the requirements of 'Mendelian randomisation', and to be confined to human leucocyte antigen (HLA)-matched sibling donors. Using different study types enabled the evaluation in view of lacking RCTs. It should be noted that these study data were generated more than 10 years ago and may not be applicable to the current medical care status.

Example 3: Localised prostate cancer. In this systematic review, we assumed a long follow-up and rare events concerning death: PCa: OS in [figure 3](#).²⁵ Localised prostate cancer as opposed to advanced prostate cancer is believed to be associated with a very good overall survival

regardless of the intervention. While invasive interventions may not improve overall survival, they may impair the health-related quality of life considerably. For example, radical prostatectomy may promise to completely remove the malignant tumour but may also disrupt erectile function in a considerable proportion of patients. According to the algorithm, a cohort study would be appropriate to estimate long-term overall survival. Concerning patient-reported outcomes such as discomfort, disability and dissatisfaction, we assume that we have short follow-up and frequent events: PCa: HRQL in [figure 3](#). According to the algorithm, an RCT would be the best choice to evaluate the patient-reported outcomes. As a single RCT was available, data on discomfort, disability and dissatisfaction were sparse and the inclusion of CCTs expanded the results considerably.²⁶ Evaluating overall survival needed a different approach than evaluating patient-reported outcomes. The obvious reluctance of patients and physicians alike to participate in RCTs corroborated the consideration of other study designs, though, restrictive inclusion criteria were necessary to enable a minimal level of quality.

Table 6 Short versus long follow-up depending on the type of disease and intervention/exposure

Diagnosis and intervention	Follow-up		Reference
	Short (early)	Long (late)	
Shortening the duration and reducing the severity of the common cold treated by vitamin C	<3 days	≥1 week	39
Early vs late radiation morbidity	<30 days	≥30 days	40
Cancer-specific survival after recurrence in patients with recurrent renal cell carcinoma	<5 years	≥5 years	41
Early or late diagnosis on patient survival in gastric cancer	<3 years	≥3 years	42
Early or late mortality after isolated first coronary bypass surgery in multivessel disease in patients with diabetes	<30 days	≥30 days	43
Early or late major adverse cardiac events after percutaneous coronary intervention in cardiac patients	<6 months	≥6 months	44

Example 4: Negative pressure wound therapy. In this systematic review, we assumed a short follow-up and frequent events concerning complete wound closure: NPWT: Closure in figure 3 and we assumed a long follow-up and rare events concerning the outcome of severe adverse events NPWT: AE in figure 3.²⁷ According to the algorithm, an RCT or a CCT would have been appropriate to evaluate the successful treatment of the disease. In 2009, the US Food and Drug Administration issued a report on six deaths and 77 other complications that were reported within a 2-year period in connection with NPWT.⁵⁰ Many of the deaths occurred in outpatient care or care homes and were caused by bleeding complications. The consideration of registry analyses and case reports were very helpful to draw attention to possible dangerous and life-threatening events.

DISCUSSION

Objective 1

In a separate paper, we concluded that “the integration of multiple study designs in systematic reviews is required and that the risk of presenting uncertain results without knowing for sure the direction and magnitude of the effect holds true for both nonrandomized and randomized controlled trials”.¹⁴ Our results appear

to be in agreement with other authors. A Cochrane review compared RCTs versus historically or concurrently controlled non-randomised trials in 2007.⁵¹ The authors concluded that, on average, the non-randomised controlled trials tend to result in larger estimates of effect than RCTs. The latest update of this Cochrane review in 2011 amended the research question and compared RCTs versus concurrently controlled non-randomised trials and excluded historically controlled ones.⁵² The authors concluded that “the results of randomized and non-randomized controlled trials sometimes differed”, namely, “in some instances nonrandomized studies yielded larger estimates of effect and in other instances randomized trials yielded larger estimates of effect”. It appears that the early firm statement expressing larger estimates in the non-randomised controlled trials changed to a less decided message.

Objective 2

We reported our experience gained during the conduct of four of our systematic reviews. These systematic reviews required the inclusion of multiple study designs to accomplish the planned evaluation of healthcare interventions. They present a few selected topics. Experiences or conclusions derived from these papers are far from being representative and not predestined to

Table 7 Rare versus frequent events depending on the type of disease and intervention/exposure

Diagnosis and intervention	Event		Reference
	Rare	Frequent	
Dying from lung cancer: lifelong non-smokers vs current smokers (≥25 cigarettes per day)	17/100 000/year	415/100 000/year	45
Prevalence of chronic obstructive pulmonary disease in 2118 lifelong never-smokers without vs with exposure to environmental tobacco smoke (ever at home and at both previous and current work)	4.2%	14.7%	46
Maternal mortality goal of the Healthy People objective in 2000	<3.3/100 000 live births	≥3.3/100 000 live births	47
Relative risk of lung cancer in current smokers vs non-smoker	Not applicable	24.0	45
Relative risk of lung cancer in non-smoker exposed vs not exposed to environmental tobacco smoke	Not applicable	2.4	48

Table 8 Examples for outcomes depending on lengths of follow-up and frequency of events

Outcome	Short follow-up		Long follow-up	
	Rare events	Frequent events	Rare events	Frequent events
Death	Population: Hypertrophic cardiomyopathy Intervention: Physical activity Aim: Incidence of sudden death after recreational sports Practical: Ethical concerns against experimental allocation	Population: Malignancies Intervention: Unrelated and mismatched haematopoietic stem cell transplantation Aim: Incidence of death due to graft rejection Practical: Ethical concerns may arise to randomise patients	Population: Acquired severe aplastic anaemia Intervention: Long-term immunosuppressive therapy Aim: Incidence of secondary malignancies Practical: Ethical concerns against randomised allocation	Population: High-risk neuroblastoma Intervention: Retinoic acid as postconsolidation therapy after high-dose chemotherapy followed by autologous haematopoietic stem cell transplantation Aim: Event-free survival in comparable groups Practical: Randomised allocation is required to provide comparable groups
Disease	Population: Neuroblastoma Intervention: Watchful waiting Aim: Incidence of spontaneous regression Practical: Ethical concerns against experimental allocation	Population: Malignancies Intervention: Unrelated and mismatched haematopoietic stem cell transplantation Aim: Incidence of acute graft vs host disease Practical: Randomised allocation is required to provide comparable groups	Population: Malignancies Intervention: Radiotherapy Aim: Incidence of myelodysplastic syndrome Practical: Ethical concerns against randomised allocation	Population: Diabetic foot ulcer Intervention: Negative pressure wound therapy Aim: Incidence of wound closure Practical: Randomised allocation is required to provide comparable groups
Discomfort Disability Dissatisfaction	Population: Acquired severe aplastic anaemia Intervention: Matched sibling donor haematopoietic stem cell transplantation Aim: Incidence of graft failure as an early and rather unexpected serious complication only observed in the transplant group not in the non-transplant group Practical: Ethical concerns against experimental allocation	Population: Low-risk localised prostate Intervention: Radical prostatectomy Aim: Incidence of erectile dysfunction and urinary incontinence Practical: Randomised allocation provides comparable groups	Population: Advanced prostate cancer Intervention: hormonal androgen deprivation therapy Aim: Incidence of emotional distortion Practical: Randomised allocation provides comparable groups	Population: Diabetic foot ulcer Intervention: Negative pressure wound therapy Aim: Incidence of amputation Practical: Randomised allocation is required to provide comparable groups

be generalised. They were conducted by the person who is also first author of the present work. Subsequently, the inferences based on the four papers and reported in the present paper may be subjective. Thus, further research by other authors and concerning other topics is recommended.

Objective 3

We did not identify an existing algorithm or a comprehensive guidance focused on finding the appropriate research design. Therefore, we developed an algorithm, which aims to guide systematic reviewers in the reasonable inclusion of various study designs in their planned systematic reviews of healthcare interventions. The

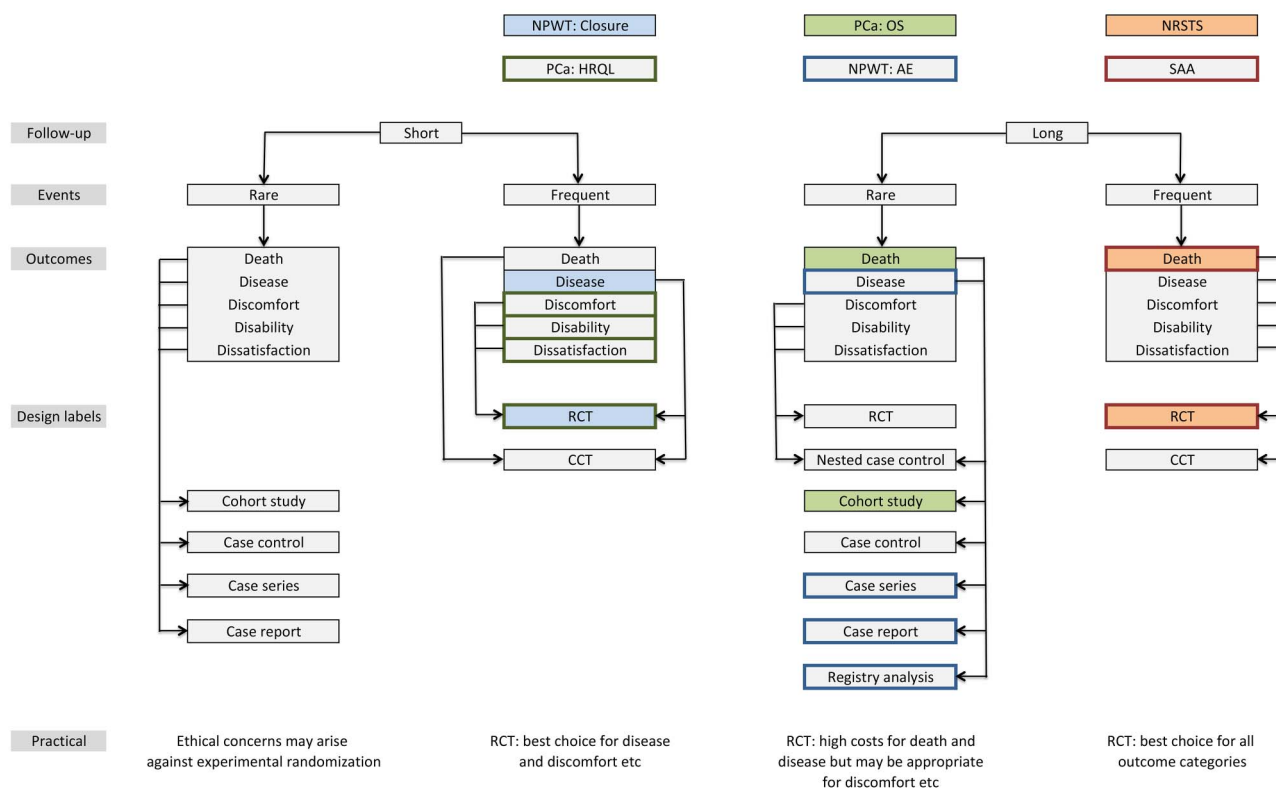
proposed algorithm cannot be applied without considering disease-specific circumstances and aims of interventions. The terms used for defining the critical decision points of the algorithm such as short versus long follow-up need to be interpreted in the context of the disease and may be unclear and not useful if used as general terms. We provided examples to show that short versus long follow-up can vary considerably depending on the disease. Similarly, we provided examples to show that the definition of rare versus frequent events has to be interpreted in the context of the type of intervention or exposure as well as the type of event. The outcomes include hard and soft outcomes, physician-reported and patient-reported outcomes and it is likely that the

Table 9 Examples for outcomes of respiratory tract diseases depending on lengths of follow-up and frequency of events

Outcome	Short follow-up		Long follow-up	
	Rare events	Frequent events	Rare events	Frequent events
Death	Viral infection may aggravate to acute myocarditis and subsequent heart failure	Lack of nourishment and lack of medicines may cause general susceptibility to life-threatening disease	Infection may affect organs such as the heart. Fibrous replacement of organ tissue may result in late arrhythmia and subsequent cardiac arrest	Lung cancer is the most common cause of cancer-related death in men and women
Disease	Bacterial infection may aggravate to community-acquired pneumonia	Infection may develop to acute sinusitis that may worsen and prolong the condition	Streptococcal pharyngitis may be complicated by chronic rheumatic heart disease	Long-term exposure to tobacco smoke is the most often cause of lung cancer
Discomfort Disability Dissatisfaction	Common cold may confine to bed and cause sick leave	Acute sinusitis may cause drowsiness, headache and sleepiness	Streptococcal pharyngitis may be complicated by rheumatic fever, which may have an involuntary movement disorder called Sydenham's chorea as a main symptom	In non-smokers, secondhand smoke may be the cause of about 20% of cases of chronic obstructive pulmonary disease, which is characterised by shortness of breath and cough

outcomes can match the purpose of the algorithm. Nevertheless, the types of outcomes have been arbitrarily chosen from a handbook of clinical epidemiology. An alternative selection of other outcomes might also be acceptable for the understanding and usefulness of the algorithm. The labelling of study designs and the descriptions of study design features are not consistently

used. Hartling 2010, while testing a tool on study design classification, reported that reviewers disagreed considerably on fundamental design characteristics, such as whether the design was experimental or observational and whether there was a control group involved or not.⁵³ Lopez-Alcade 2011 reported that “Cochrane review groups did not use common study design labels

**Figure 3** Algorithm with pathways backtracked in four completed systematic reviews.

and did not explicitly describe all study design features suggested by the Cochrane Handbook".⁵⁴ We are confident that the algorithm is a tool helping to bring seminal features of a systematic review to the attention of anyone who is planning to conduct a systematic review. It has the potential to help to reorientate oneself to major features of the studies eligible for an evaluation of a healthcare intervention. The benefit is the provision of awareness, and it is certainly not a new regulation. The intention is to provide a guide and a decision support tool that might be used fully or partially by persons who are going to prepare a systematic review. While preparing a systematic review, it may be important at an early time point to identify the relevant and the most appropriate study designs necessary to find answers for a variety of prespecified outcomes. It might also be of interest for persons who evaluate the quality of systematic reviews and might want to check whether the all study designs have been considered that should have been considered. Therefore, we think that it may reduce the time burden on review authors and contribute to the production of a higher quality review.

Objective 4

The plausibility check is a crude approach to speculate if the theory-based algorithm could be sensibly applied in practice. Thus, further research could facilitate a more objective and statistically measurable testing of the usefulness of the algorithm. It is recommended to let various systematic reviewers backtrack the algorithm independently and to apply the algorithm on more systematic reviews with different topics. We could follow and confirm the appropriateness of the pathways for all described examples of systematic reviews. In one example, the algorithm selected RCTs as the best choice but due to the lack of RCTs it was decided to rely on non-randomised studies. This example showed that it is important to build flexibility into the algorithm, which enables the systematic reviewer to extend or change the inclusion criteria to other study designs in case that certain unexpected conditions may emerge or practical concerns exist. While conducting systematic reviews, we observed the critical importance of case reports and registry analysis for the evaluation of serious adverse events. We mentioned above the Food and Drug Administration (FDA) report on adverse events after NPWT, which covered 2 years from 2007 to 2009. In a recent update, the FDA included two additional years covering a total of 4 years from 2007 to 2011.⁵⁵ The adverse events increased to 12 deaths and 174 injuries. With respect to the added cases, bleeding was again the major cause of the most serious adverse events and the majority of adverse events occurred at home or in long-term care facilities. All RCTs on NPWT were conducted in hospitals and were unable to provide this information. In France between 1998 and 2004, 21 drugs were reported to be withdrawn from the market for safety reasons. The withdrawal of 19 of 21 drugs was based on

case reports and only 1 case was supported by RCT.⁵⁶ In the European Community between 2002 and 2011, case reports contributed to the withdrawal of 18 of 19 drugs.⁵⁷

CONCLUSIONS

We are confident that the algorithm can assist to bring seminal features of a systematic review to the attention of anyone who is planning to conduct a systematic review. It aims to provide awareness and we think that it may reduce the time burden on review authors and may contribute to the production of a higher quality review.

Contributors FP and JK conceived and designed the experiments. FP performed the experiment, analysed the data and contributed analysis tools. FP and JK wrote the manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Victora CG, Habicht JP, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health* 2004;94:400–5.
2. Sackett DL, Straus SE, Richardson WS, et al. *Evidence-based medicine. How to practice and teach EBM*. 2nd edn. Edinburgh: Churchill Livingstone, 2000.
3. Clarke N. *Evidence-based medicine tutorial*. Tallahassee: Medical Informatics, College of Medicine, Florida State University, 2014. <http://fsu.edu> (access date 20 May 2015).
4. Vandembroucke JP. Observational research, randomised trials, and two views of medical science. *PLoS Med* 2008;5:e67.
5. G-BA. *Verfahrensordnung stand: 21. Februar 2013* [in German]. Berlin: Gemeinsamer Bundesausschuss (G-BA), 2013. <http://www.g-ba.de> (access date 20 May 2015).
6. CEBM. *Levels of evidence*. Oxford: Centre of Evidence-Based Medicine (CEBM), University of Oxford, 2009. <http://www.cebm.net> (access date 20 May 2015).
7. CRD. *Systematic reviews. CRD's guidance for undertaking reviews in health care*. York: Centre for Reviews and Dissemination (CRD), University of York, 2009. <http://www.york.ac.uk> (access date 20 May 2015).
8. Khan KS, Kunz R, Kleijnen J, et al. *Systematic reviews to support evidence-based medicine. How to review and apply findings of healthcare research*. London: Hodder Arnold, 2011.
9. NICE. *Guide to the methods of technology appraisal*. London: National Institute for Clinical Excellence (NICE), 2004. <http://www.nice.org.uk> (access date 20 May 2015).
10. SIGN. SIGN 50. A guideline developer's handbook. Revised November 2011. Annex B page 51. Edinburgh: Scottish Intercollegiate Guidelines Network (SIGN), 2011. <http://www.sign.ac.uk> (access date 20 May 2015).
11. Campbell PG, Malone J, Yadla S, et al. Comparison of ICD-9-based, retrospective, and prospective assessments of perioperative complications: assessment of accuracy in reporting. *J Neurosurg Spine* 2011;14:16–22.
12. CEBM. *Study designs*. Oxford: Centre of Evidence-Based Medicine (CEBM), University of Oxford, 2014. <http://www.cebm.net> (access date 20 May 2015).

13. NICE. *Guide to the methods of technology appraisal 2013*. London: National Institute for Health and Care Excellence (NICE), 2013. <http://www.nice.org.uk> (access date 20 May 2015).
14. Peinemann F, Tushabe DA, Kleijnen J. Using multiple types of studies in systematic reviews of health care interventions—a systematic review. *PLoS ONE* 2013;8:e85035.
15. Fletcher RH, Fletcher SW. *Clinical epidemiology. The essentials*. 4th edn. Philadelphia: Lippincott Williams and Wilkins, 2005.
16. Reeves BC, Deeks JJ, Higgins JPT, et al. Chapter 13. Including non-randomized studies. *Cochrane handbook for systematic reviews of interventions version 5.1.0* (updated March 2011). Chichester: The Cochrane Collaboration, 2011. <http://www.cochrane-handbook.org> (access date 20 May 2015).
17. Cochrane Community (beta). *Glossary*. The Cochrane Collaboration. <http://community.cochrane.org> (access date 20 May 2015).
18. EBMT. *Guidelines for the conduct of registry-based studies using the EBMT database (version 2), last modified: 26/04/11*. Leiden: European Society for Blood and Marrow Transplantation (EBMT). <http://www.ebmt.org> (access date 20 May 2015).
19. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011;64:395–400.
20. Matchar DB. Introduction to the methods guide for medical test reviews. In: Chang SM, Matchar DB, eds. *Methods guide for medical test reviews*. Rockville: Agency for Healthcare Research and Quality (AHRQ) 2012.
21. White CM, Ip S, McPheeters M, et al. Using existing systematic reviews to replace de novo processes in conducting Comparative Effectiveness Reviews. In: AHRQ. *Methods guide for comparative effectiveness reviews*. Rockville: Agency for Healthcare Research and Quality (AHRQ), 2009.
22. Peinemann F, Smith LA, Bartel C. Autologous hematopoietic stem cell transplantation following high dose chemotherapy for non-rhabdomyosarcoma soft tissue sarcomas. *Cochrane Database Syst Rev* 2013;(8):CD008216.
23. Peinemann F, Smith LA, Kromp M, et al. Autologous hematopoietic stem cell transplantation following high-dose chemotherapy for non-rhabdomyosarcoma soft tissue sarcomas. *Cochrane Database Syst Rev* 2011;(2):CD008216.
24. Peinemann F, Bartel C, Grouven U. First-line allogeneic hematopoietic stem cell transplantation of HLA-matched sibling donors compared with first-line ciclosporin and/or antithymocyte or antilymphocyte globulin for acquired severe aplastic anemia. *Cochrane Database Syst Rev* 2013;(7):CD006407.
25. Peinemann F, Grouven U, Bartel C, et al. Permanent interstitial low-dose-rate brachytherapy for patients with localised prostate cancer: a systematic review of randomised and nonrandomised controlled clinical trials. *Eur Urol* 2011;60:881–93.
26. Peinemann F, Grouven U, Hemkens LG, et al. Low-dose rate brachytherapy for men with localized prostate cancer. *Cochrane Database Syst Rev* 2011;(7):CD008871.
27. Peinemann F, Sauerland S. Negative-pressure wound therapy: systematic review of randomized controlled trials. *Dtsch Arztebl Int* 2011;108:381–9.
28. AHRQ. *Negative pressure wound therapy devices*. Rockville: Agency for Healthcare Research and Quality (AHRQ), 2009. <http://archive.ahrq.gov> (access date 20 May 2015).
29. ASERNIP-S. *General guidelines for assessing, approving & introducing new surgical procedures into a hospital or health service*. Stepney: Australian Safety and Efficacy Register of New Interventional Procedures—Surgical (ASERNIP-S), Royal Australasian College of Surgeons (RACS), 2009. <http://www.surgeons.org> (access date 20 May 2015).
30. CADTH. *Guidelines for authors of CADTH health technology assessment reports*. Ottawa: Canadian Agency for Drugs and Technologies in Health (CADTH), 2003. <https://www.cadth.ca> (access date 20 May 2015).
31. HAS. *General method for assessing health technologies*. Paris: Haute Autorité de Santé (HAS), French National Authority for Health, 2007. <http://www.has-sante.fr> (access date 20 May 2015).
32. IQWiG. *Methoden Version 4.1 [in German]*. Cologne: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) [German Institute for Quality and Efficiency in Health Care], 2013. <https://www.iqwig.de> (access date 20 May 2015).
33. MRC. *Developing and evaluating complex interventions: new guidance*. London: Medical Research Council (MRC), 2008. <http://www.mrc.ac.uk> (access date 20 May 2015).
34. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215–8.
35. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:1–173.
36. Egger M, Davey Smith G, Altman DG. *Systematic reviews in health care. Meta-analysis in context*. 2nd edn. London: B.M.J. Publishing Group, 2006.
37. Guyatt GH, Rennie D, eds. *Users' guides to the medical literature. A manual for evidence-based clinical practice*. 5th edn. Chicago: AMA Press, 2005.
38. Reeves BC, van Binsbergen J, van Weel C. Systematic reviews incorporating evidence from nonrandomized study designs: reasons for caution when estimating health effects. *Eur J Clin Nutr* 2005;59 (Suppl 1):S155–61.
39. Hemila H, Chalker E. Vitamin C for preventing and treating the common cold. *Cochrane Database Syst Rev* 2013;(1):CD000980.
40. RTOG. *Guidance for routine adverse event reporting on RTOG protocols (initial 4.11.2013)*. Philadelphia: Radiation Therapy Oncology Group (RTOG), 2013. <http://www.rtog.org> (access date 20 May 2015).
41. Brookman-May SD, May M, Shariat SF, et al. Time to recurrence is a significant predictor of cancer-specific survival after recurrence in patients with recurrent renal cell carcinoma—results from a comprehensive multi-centre database (CORONA/SATURN-Project). *BJU Int* 2013;112:909–16.
42. Arvanitakis C, Nikopoulos A, Giannoulis E, et al. The impact of early or late diagnosis on patient survival in gastric cancer in Greece. *Hepatogastroenterology* 1992;39:355–7.
43. Calafiore AM, Di Mauro M, Di Giammarco G, et al. Effect of diabetes on early and late survival after isolated first coronary bypass surgery in multivessel disease. *J Thorac Cardiovasc Surg* 2003;125:144–54.
44. Pedersen SS, Martens EJ, Denollet J, et al. Poor health-related quality of life is a predictor of early, but not late, cardiac events after percutaneous coronary intervention. *Psychosomatics* 2007;48:331–7.
45. Doll R, Peto R, Boreham J, et al. Mortality from cancer in relation to smoking: 50 years observations on British doctors. *Br J Cancer* 2005;92:426–9.
46. Hagstad S, Bjerg A, Ekerljung L, et al. Passive smoking exposure is associated with increased risk of COPD in never smokers. *Chest* 2014;145:1298–304.
47. NCHS. *Healthy people 2000 final review*. Hyattsville: National Center for Health Statistics (NCHS), Public Health Service, 2001. <http://www.cdc.gov> (access date 20 May 2015).
48. Hackshaw AK, Law MR, Wald NJ. The accumulated evidence on lung cancer and environmental tobacco smoke. *BMJ* 1997;315:980–8.
49. ORDR. *What is a rare disease?* Bethesda: Office of Rare Diseases Research (ORDR), National Institutes of Health (NIH), 2014. <http://rare diseases.info.nih.gov> (access date 20 May 2015).
50. FDA. *Preliminary public health notification: serious complications associated with negative pressure wound therapy systems*. Date issued: November 13, 2009. Silver Spring: U.S. Food and Drug Administration (FDA), 2009. <http://www.fda.gov> (access date 20 May 2015).
51. Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2007;(2):MR000012.
52. Odgaard-Jensen J, Vist GE, Timmer A, et al. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev* 2011;(4):MR000012.
53. Hartling L, Bond K, Harvey K, et al. Developing and Testing a Tool for the Classification of Study Designs in Systematic Reviews of Interventions and Exposures. Agency for Healthcare Research and Quality; December 2010. Methods Research Report. AHRQ Publication No. 11-EHC-007. <http://effectivehealthcare.ahrq.gov> (access date 20 May 2015).
54. Lopez-Alcade J, Calcerrada N, Callejo D, et al. Defining study designs in The Cochrane Collaboration: do all Cochrane Review Groups talk the same language? Poster presentation at the 19th Cochrane Colloquium. Madrid, Spain [abstract], 2011:19–22. <http://bvssalud.org> (access date 20 May 2015).
55. FDA. *Update on serious complications associated with negative pressure wound therapy systems*. Date issued: February 24, 2011. Silver Spring: U.S. Food and Drug Administration (FDA), 2011. <http://www.fda.gov> (access date 20 May 2015).
56. Olivier P, Montastruc JL. The nature of the scientific evidence leading to drug withdrawals for pharmacovigilance reasons in France. *Pharmacoepidemiol Drug Saf* 2006;15:808–12.
57. McNaughton R, Huet G, Shakir S. An investigation into drug products withdrawn from the EU market between 2002 and 2011 for safety reasons and the evidence used to support the decision-making. *BMJ Open* 2014;4:e004221.