

RESEARCH

Classifying information-sharing methods: a citation-mining systematic review

Georgios F. Nikolaidis^{*}, Beth Woods, Stephen Palmer and Marta O. Soares

Abstract

Background: Sparse relative effectiveness evidence is a frequent problem in Health Technology Assessment (HTA). Where evidence directly pertaining to the decision problem is sparse, it may be feasible to expand the evidence-base to include studies that relate to the decision problem only indirectly: for instance, when there is no evidence on a comparator, evidence on other treatments of the same molecular class could be used; similarly, a decision on children may borrow-strength from evidence on adults. Usually, in HTA, such indirect evidence is either included by ignoring any differences ('lumping') or not included at all ('splitting'). However, a range of more sophisticated methods exists, primarily in the biostatistics literature. The objective of this study is identify and classify the breadth of the available information-sharing methods.

Methods: Forwards and backwards citation-mining techniques were used on a set of seminal papers on the topic of information-sharing. Papers were included if they specified (network) meta-analytic methods for combining information from distinct populations, interventions, outcomes or study-designs.

Results: Overall, 89 papers were included. A plethora of evidence synthesis methods have been used for information-sharing. Most papers ($n = 78$) described methods that shared information on relative treatment effects. Amongst these, there was a strong emphasis on methods for information-sharing across multiple outcomes ($n = 39$) and treatments ($n = 23$), with fewer papers focusing on study-designs ($n = 10$) or populations ($n = 6$). We categorise and discuss the methods under four 'core' relationships of information-sharing: functional, exchangeability-based, prior-based and multivariate relationships, and explain the assumptions made within each of these core approaches.

Conclusions: This study highlights the range of information-sharing methods available. These methods often impose more moderate assumptions than lumping or splitting. Hence, the degree of information-sharing that they impose could potentially be considered more appropriate. Our identification of four 'core' methods of information-sharing allows for an improved understanding of the assumptions underpinning the different methods. Further research is required to understand how the methods differ in terms of the strength of sharing they impose and the implications of this for health care decisions.

Abstract wordcount: 325 words

Manuscript wordcount: 4615 words

Keywords: Meta-analysis; Network meta-analysis; Borrowing-strength; Indirect evidence; Information-sharing

Background

Health Technology Assessment (HTA) is the systematic evaluation of the properties, effects and impact of health technologies with a view to inform decision-making in health care [1]. Regardless of whether or not a system functions under explicit budget constraints, resources spent could have always been used for alter-

native purposes. Therefore, policy-makers are always faced with difficult decisions about whether interventions should be funded. This requires an assessment of whether the benefits of an intervention are sufficient to justify the health opportunity costs of funding it [2]. It follows that a set of tools ought to be used so that policy-makers can rationally and transparently decide about the adoption of a given health technology [3].

Decision analysis provides a quantitative framework that brings together all relevant evidence on the im-

^{*}Correspondence: georgios.nikolaidis@york.ac.uk

The University of York, Centre for Health Economics, Alcuin A Block, Heslington, YO10 5DD York, UK

Full list of author information is available at the end of the article

pect of an intervention on health outcomes and costs, whilst making explicit judgements about how different types and sources of evidence are linked together (model structure) and which elements are relevant to decision-making (reflecting social values). The outputs of a Decision Analytic Model (DAM) include incremental costs and benefits and can be useful for decision-makers [4].

Each input within a DAM is a parameter and constitutes a potential research question that can be informed by evidence which is typically identified using literature reviews. To assist study selection when identifying evidence for reviews, research questions are defined using the PICOS framework, where P stands for Population, I for Intervention, C for Comparator, O for Outcome, and S for Study-design [5]. Typically, reviewers exclude studies deviating from any of the PICOS criteria; that is, they usually only include studies providing *direct evidence*. Hence, direct evidence on relative effectiveness comprises of one or more randomised studies, evaluating the intervention(s) under assessment, recruiting patients from the population of interest, and measuring effects on all relevant outcomes.

Where multiple studies exist to inform the same parameter, these can be synthesised to generate a single estimate that represents the evidence-base. However, due to the availability of evidence, the need for evidence synthesis is more common for parameters like the Relative Treatment Effect (RTE) than for other DAM inputs such as costs and Quality of Life (QoL). Standard Meta-Analysis (MA) and Network Meta-Analysis (NMA) methods [6, 7] are commonly used to synthesise the evidence base and provide DAMs with relative effectiveness inputs.

However, in HTA, direct evidence may be sparse, heterogeneous, or limited in other ways and synthesis may become problematic. Where evidence is sparse, it may not be possible to obtain the required RTE estimates, and even when they can be obtained, they may be highly uncertain and may not be robust due to assumptions imposed in the analysis [8, 9]. Evidence sparsity may also prevent appropriate exploration of heterogeneity because small studies are at higher risk of enrolling unrepresentative populations [10].

A policy relevant alternative to limited or sparse data may be to extend the evidence base beyond the direct evidence. A topical example concerns paediatric indications for which the evidence-base is typically sparse due to the regulatory restrictions on trials. To support decision-making for this population, the U.S. Food and Drugs Administration (FDA) [11] and the European Medicines Agency (EMA) now propose that “*The evidence needed to address the research questions*

that are important for marketing authorisation of a given product in the target population might be modified based on what is known for other population” [12]. Whilst in the aforementioned example the evidence is extended to consider another population, in principle, indirect evidence may relate to any other level of PICOS (Figure 1)- it may include studies assessing a different, but related, treatment or pertaining to a different study-design than what is specified in the research question. Note that, in this context, NMAs also consider indirect evidence, pertaining to other treatment comparisons, to inform the treatment effect(s) of primary interest [13].

[INSERT FIGURE 1 HERE]

Within a decision-making context, the use of indirect evidence, as long as it is judged relevant, contributes to accountability by allowing for all relevant evidence to be considered. Combining all relevant sources of evidence may yield more precise estimates than the direct evidence alone and allow better characterisation of heterogeneity and uncertainty. However, when indirect evidence are not sufficiently relevant or of high-quality, using indirect evidence may also introduce bias and inflate heterogeneity estimates.

The use of indirect evidence to support decision-making is not exclusive to the aforementioned regulatory context and has permeated HTA processes. Examples can be found in Technology Appraisals (TAs) conducted by the National Institute for Health and Care Excellence (NICE) to inform routine use of technologies in the National Health Service (NHS) in England and Wales. For instance, TA445 [14] considered adult studies to complement a sparse paediatric evidence base. Also, relative effectiveness has been generalised across subgroups of different Hepatitis C genotypes [15]. These two examples use indirect evidence by considering both sources perfectly generalisable (‘lumping’), as an alternative to being considered completely independent (‘splitting’). There are, however, examples of appraisals which use indirect evidence in more sophisticated ways. For instance, (TA383) [16] used indirect evidence across interventions by assuming a ‘class-effect’ between treatments that function through the same molecular pathway. TA139 [17] and TA168 [18] simultaneously modelled two outcomes leveraging their correlation structure and TA244 [19] modelled a network of interventions with multiple treatment components assuming that the relative effect of an intervention is the sum of the relative effects of its comprising components.

Inevitably, a judgement on whether the indirect evidence is relevant is always required. However, what is often not made explicit is that, where both direct and

indirect evidence are considered, there should be appropriate consideration for the extent of information-sharing permitted by different synthesis methods (i.e. the extent to which the indirect evidence is allowed to affect the estimates obtained by using only the direct evidence).

The objective of this review is to identify information-sharing evidence synthesis methods that have been used in the literature and improve understanding of these methods by making explicit the fundamental assumptions underpinning them. We do so by identifying the ‘core’ relationships used to share information. This review increases awareness around the breadth of available information-sharing methods and aids transparency in information-sharing methods choice. To our knowledge, this topic has not been explored in the past with a clear policy focus.

Methods

Given the lack of consistent terminology in the literature referring to methods that combine direct and indirect evidence, keyword-based search methods [20] were not used. Instead, ‘citation-mining’ methods [21] were used, which are efficient [22] and have been used for similar reviews [23]. The review protocol was developed by the authors and validated by the first author’s Doctoral Thesis Advisory Panel. The process consisted of the following steps: initially, a list of seminal/influential papers was compiled after consulting with experts and conducting a scoping literature review. The scoping review included terms to represent a variety of fields including MA, NMA, multi-parameter evidence synthesis, and the incorporation of evidence on historical controls in trial-design. Although the last category is outside the scope of this work, it may have influenced the extension of methods to the MA/NMA field.

The citations of 7 seminal papers [8, 24–29] were searched in the Web of Science (WoS) on 20-Feb-2019. Subsequently, articles were identified that cited -forwards citation-mining- or were cited by the seminal papers -backwards citation-mining-. Articles were included if they mathematically specified MA or NMA models that combine information pertaining to multiple populations, interventions, outcomes, study-designs or utilise evidence from an external source such as previous meta-analyses. Importantly, papers that used only standard NMA methods were excluded even though they share information across treatment comparisons, because such methods are well established in the literature. Further information on the search strategy and inclusion and exclusion criteria is provided in Additional file 1.

From each included paper, the synthesis model was isolated and from within it, methods facilitating information-sharing were extracted. Methods were subsequently categorised according to the ‘core’ relationship that they used to enable information-sharing. When papers tackled multiple synthesis challenges simultaneously (e.g. [30, 31]), the issues they dealt with were isolated along with the method used to address each. The PICOS level of indirectness was also extracted. The search was conducted in Zotero version 5.0.69. The PRISMA checklist for systematic reviews is provided in Additional file 2.

Results

Characteristics of the included studies

The review identified 89 papers (Figure 2). The majority ($n = 78$) described methods that shared information on relative treatment effects. Other studies used methods to share information on comparison-specific meta-regression slopes ($n = 4$), comparison-specific between-studies heterogeneities ($n = 6$), or study-specific baselines ($n = 2$). Overall, there was a balance amongst papers that developed methods within MA ($n = 45$) and NMA ($n = 44$). There was a strong emphasis on methods for information-sharing across multiple outcomes (39 papers) and treatments (23 papers), with fewer papers focusing on study-designs (10 papers) or populations (6 papers) (Table 1). Note that some papers described methods sharing information on several types of parameters and across more than one PICOS level (e.g. [32]). A full list of the included papers along with a description of how information was shared within each paper can be found in Additional file 3.

[INSERT FIGURE 2 HERE]

We also identified the most common synthesis challenges addressed by the included papers. Amongst papers sharing information across populations, 3 [14, 33, 34] developed models to accommodate simultaneous synthesis of adult and paediatric evidence, 1 [35] described models that allowed information-sharing between patients subgroups, and 2 [36, 37] provided model extensions for baseline risk adjustment. Amongst papers sharing information across interventions, 7 [38–44] simultaneously synthesised multiple dosages of the same treatment, 7 [32, 35, 40, 44–47] shared across interventions that fall under the same ‘class’, and 5 [30, 45, 48–50] dealt with complex interventions (i.e. treatments that comprise from multiple components). Amongst papers sharing information across study-designs, 4 [51–54] combined randomised and non-randomised studies and 12 [51, 52, 55–64] dealt with studies’ internal or external biases.

Amongst papers that sharing information across outcomes, 2 [31, 65] considered structurally related outcomes (for instance, when one outcome has to occur before the other), 5 [43, 66–69] combined evidence from studies that reported at multiple/different follow-up periods, and 22 papers [26, 30, 31, 65, 70–87] correlated outcomes. Finally, 5 papers [24, 58, 88–90] described how evidence from previous meta-analyses (meta-epidemiological evidence) can be incorporated in analyses.

‘Core’ relationships for information-sharing

The methods identified were classified according to the ‘core’ relationship facilitating information-sharing. Four ‘core’ methods were identified (Figure 3) : 1) **functional** relationships which include deterministic functions among model parameters resulting in a reduced number of parameters that need to be estimated; 2) **exchangeability-based** relationships which assume that a set of parameters are drawn from a common distribution that allows them to be shrunk towards its mean; 3) **prior-based** relationships which employ a Bayesian framework to ‘load’ the indirect evidence in prior distributions and 4) **multi-variate** methods which assume that model parameters are correlated and enable information-sharing through the covariance structure.

[INSERT FIGURE 3 HERE]

Table 1 classifies papers according to the ‘core’ method used and the PICOS level of indirectness. It shows that some ‘core’ relationships are preferred when information is shared across specific PICOS levels. For instance, most of the identified papers sharing across interventions either use functional or exchangeability-based relationships, and no example using priors was found. Also, papers that use multivariate relationships, do so to share information across related outcomes, not across populations or study-designs. This may be partly because the information required to implement multivariate methods that capture the between-study correlation for multiple populations or study-designs (i.e. studies reporting separately for two different populations or designs) is usually unavailable in the literature because single studies rarely enroll or report separately for multiple populations and cannot, by definition, pertain to multiple designs.

Functional relationships

The simplest functional relationship is lumping (i.e. common effects) where all data points inform a single parameter independently of whether the evidence is direct or indirect. Examples include pooling RTEs across time-points [32] or (sub-)populations [14, 35] as well as pooling between-trial heterogeneity parameters [92] or meta-regression slopes [91].

Another type of functional relationship is constraint where a strict inequality is imposed among parameters. In a Bayesian framework, information-sharing is facilitated by preventing simulation samples that do not conform to the specified constraint. Such methods have been used to relate RTEs across dosages, expressing that higher dosages are expected to exhibit larger RTE [39, 40], describe structurally-related outcomes [65] and specify second-order consistency equations that impose a triangle inequality on the comparison-specific between-trial variances [94, 95].

Meta-regression-type methods have also been suggested. In the examples found, the relationships were usually linear -on the modelling scale- with one RTE component independent and another RTE component dependent on a particular study characteristic. The most common example in this category is bias-adjustment methods, primarily used to synthesise studies of different designs. Bias-adjustment methods broadly fall into two categories: general frameworks that adjust the RTE for biases affecting internal and external validity provided that the extent of bias can be either estimated from empirical evidence or elicited from experts [55, 56, 63, 64], and approaches that adjust for bias due to particular study-level characteristics (considered proxies for study quality such as their size [47, 59–62], publication year [104], or risk-of-bias [57, 58]). Meta-regression-type relationships have also been used for complex interventions. In their simplest form, they model the RTE of a complex intervention as the sum of RTEs of its treatment components [30, 45, 49, 50]. More sophisticated approaches allow for synergistic or antagonistic relationships by suggesting functions that also contain treatment interaction RTE components [48]. Other applications include approaches that model the RTEs measured in two survival outcomes (e.g. time-to-mortality and time-to-progression) by assuming that they only differ by a constant component which is invariant across treatment comparisons [31], models that assume a linear relationship between dosage and RTE [44], and methods for baseline-risk adjustment [36].

Finally, more complex, non-linear, relationships have also been presented in the literature, namely those enabling the synthesis of RTEs across a range of dosages using the Emax model [38, 41, 42] commonly employed in pharmacokinetics and those enabling the sharing of information across follow-up periods [66, 67].

Exchangeability-based relationships

The simplest exchangeability-based relationship uses a random effect to relate a set of parameters; in this way accounting for heterogeneity without explicitly modelling its source(s). The random effect assumes

Table 1 A categorisation of papers that share information on the relative effectiveness parameter according to the ‘core’ relationship that they use and the PICOS level that direct and indirect evidence differ in.

PICOS level of indirectness	‘Core’ relationship			
	Functional	Exchangeability-based	Prior-based	Multivariate
Interventions	[30, 32, 35, 38–45, 45, 48–50, 58, 59, 91–95]	[32, 35, 36, 39, 40, 43–47, 59, 91, 94, 95]	No references	[45]
Populations	[14, 35]	[34, 36, 37]	[33, 34, 36]	No references
Outcomes	[32, 65–67]	[43, 66]	No references	[26, 30, 31, 65, 68, 69, 71–87, 96–103]
Designs	[47, 55–64, 104]	[51, 52, 54, 105]	[48, 51–53, 105–107]	No references

that all parameters are drawn from a distribution, implying that individual parameters are shrunken towards the random effect mean; this can happen to a greater or lesser extent, depending on the precision and discrepancy of each individual estimate in relation to the random effect mean. Examples of parameters to which random-effects have been applied include: RTEs of different dosages of the same treatment [39], comparison-specific meta-regression slopes [36, 47, 59, 91], comparison-specific between-trial variances [94, 95], and study-specific baseline-risks [36, 37].

Random-walks are another form of exchangeability relationship. They assume that data points which are more similar with respect to a particular characteristic are expected to exhibit more similar RTEs. Examples include approaches assuming that the RTE of a particular dosage or follow-up period is drawn from a distribution centred around the RTE of its adjacently lower or higher dosage [39] or follow-up period [43, 66].

Multi-level models also use exchangeability, but apply it to the hierarchical/clustered structure of the available data. As such, exchangeability is applied at a first level within specific groups of parameters (i.e. multiple random effects are applied, each within groups of RTEs from studies showing a particular characteristic) and at a second level across the group-specific hyper-parameters. This is shown in Figure 4, where in the bottom level, studies are categorised according to a characteristic and a different random effect is imposed within every category, producing group-specific basic parameters and heterogeneities. Subsequently, in the top-level, exchangeability is also assumed across the group-specific basic parameters which are shrunken towards an overall, global, group-independent, hyper-mean. Examples include ‘class-effects’ models where, on top of the classical Random-Effects (RE) NMA models, the basic parameters of treatments that function through the same mechanism are assumed to be drawn from a common distribution with an overall ‘class’ mean and an across-treatments, within-class, heterogeneity [32, 35, 40, 44–46]. Class-effect approaches have also been imposed across comparison-specific meta-regression slopes [47, 91]. Multi-level models have been

suggested to combine adult and paediatric evidence [34] and studies of different designs [51, 52, 54, 105].

[INSERT FIGURE 4]

Prior-based relationships

Direct and indirect evidence can also be combined through the use of prior distributions. The process usually consists of two-steps where initially the indirect evidence is analysed and subsequently the resulting distribution is used as a prior in the analysis of the direct evidence. Of note is that this approach is mathematically equivalent to lumping, which was described under functional relationships. Examples include the combination of adult and paediatric evidence [34] or randomised and non-randomised evidence [51, 52, 105, 107]. The prior can additionally be adjusted for bias or its precision decreased [51]. Alternative ways to define the prior include the use of meta-epidemiological evidence or expert elicitation. The former has been used primarily for bias-adjustment [48], whilst both the former [24, 88, 89] and the latter [108] have been used to define prior a distribution for the between-trials heterogeneity.

More nuanced prior-based approaches such as mixtures of priors have also been used. Here, the informative prior (distribution representing the indirect evidence) is not used at face value, but instead mixed with a vague prior according to weights that may be specified by the analyst or estimated within the synthesis model. The resulting informative prior is typically heavy-tailed, and allows for ‘adaptive’ information-sharing whereby information-sharing is stronger when the direct and indirect evidence are in agreement and weaker when they conflict [33]. Mixtures of priors have been used to combine evidence on RTE and between-studies heterogeneity across adults and children [33] and to analyse the study-specific baseline parameters from studies that enroll populations with different baseline risk [36]. The use of mixtures of priors has also been discussed for the synthesis of randomised and non-randomised evidence [51].

Finally, a flexible method that has been proposed is the power-prior [109]. In this method, the likelihood of the indirect evidence is raised to a power scalar

$0 \leq a \leq 1$ which reflects the perceived similarity between the two sources of evidence. When $a = 1$ the results are equivalent to ‘lumping’ and when $a = 0$ results are identical to ‘splitting’. The power parameter, a , needs to be specified, and it has been proposed to be elicited [110] or varied in sensitivity analysis [111]. Power priors have been used to combine observational and randomised evidence [53] and for the synthesis of adult and paediatric evidence [34].

Multivariate relationships

Multi-variate relationships have primarily been used to share information across multiple outcomes. Multivariate meta-analysis correlates the various outcomes and may separate within- and between-studies correlations [72]. At the within-study level, the study-specific correlations arise due to differences among the included patients and indicate how the outcomes co-vary across individuals within the study. For example, patients who, due to a baseline characteristic that makes their disease more severe, show high values for outcome A, are also more likely to yield high values for outcome B. At the between-studies level, correlations arise mainly due to study-level differences such as the distribution of the patient-level characteristics across studies. For instance, studies that enroll more severe cases and therefore may show high values for the mean of outcome A, are also more likely to result in high values for the mean of outcome B, whilst studies enrolling less severe cases may show lower mean values for both outcomes.

Multivariate methods have been developed to consider two [73, 74, 84], three, or more correlated outcomes [26, 87], accommodate the simultaneous analyses of multiple treatments [75, 77], and assess the relationship between surrogate and final outcomes [80, 96]. Given that within-trial correlations are commonly not reported, authors have suggested the use of external data to inform these parameters [71] or, when external data is not available, methods that approximate the within-study co-variances [86]. Further extensions have been developed to handle missing data [82] and allow modelling of heterogeneity and inconsistency using two different variance components [81].

To accommodate cases where the within-trials correlations are unavailable and cannot be otherwise obtained, alternative methods, which require the same data as a univariate approach and do not separate within- and between-trials correlations have been suggested for MA [79, 112] and NMA [75]. Assuming that the overall correlation is not very strong, these methods perform very similarly to their counterpart, which separates the two correlations, whilst preserving their benefits against the univariate approach.

Finally, some methods only account for either the within- or the between- studies correlations. For example, to model mutually exclusive outcomes, it has been suggested to only account for the within-trials negative correlations which are induced by the competing risks structure of the data (i.e. the more patients that reach an outcome, the fewer the patients that reach another outcome) [76]. Also, other approaches have only modelled the between-studies covariance matrix to allow simultaneous synthesis of multiple outcomes [30, 31, 65, 78], accommodate outcomes reported at several follow-up periods [68, 69] and enable information-sharing across different treatment components of complex interventions [45].

Discussion

The aim of this review was to identify and classify evidence synthesis methods that have been used to combine evidence from sources that relate directly and indirectly to a particular research question and classify relationships. A wide range of methods have been developed to share information between populations, treatments, outcomes and study-designs. We found that across the breadth of methods identified, four ‘core’ relationships are used to facilitate information-sharing. These are functional, exchangeability-based, prior-based, and multivariate relationships (Figure 3).

This review highlights the breadth of methods options that can facilitate information-sharing. Although, typically, particular relationships are used preferentially to share information on specific information-sharing contexts, it is likely that several methods are applicable and analysts would need to choose which method is more appropriate. This paper highlights that appropriate considerations need to be made when choosing ‘core’ relationships and methods as choices are likely to influence the degree of information-sharing. Specifically, method selection may be informed by the following considerations; the first is the plausibility of the assumptions imposed by the methods in the context of interest. By classifying methods according to the ‘core’ relationship that enables information-sharing, we hope to facilitate a clearer discussion about the plausibility of these assumptions in the decision context of interest.

The second is the degree of information-sharing that is imposed between direct and indirect evidence. Within the literature, there is limited exploration of how much different methods borrow-strength from indirect evidence, though for multivariate methods, it has been noted that information-sharing is ‘usually modest’ [26, 97] and, sometimes, instead of ‘borrowing-strength’, multi-variate methods may end up ‘borrowing-weakness’ [113]. The few studies that

have assessed the degree of information-sharing typically consider only the degree of precision gains [114] rather than also examining how the point estimate - which is also important for decision making - changes. Further research to understand the extent to which different methods share information is warranted.

Finally, decision-makers may be interested in exploring different levels of information-sharing. One way to do that is by using prior-based methods that allow some control on the degree of information-sharing. For instance, an informative prior may use either the posterior distribution of the mean, or the predictive distribution of the indirect evidence. The former is equivalent to lumping, whilst the latter imposes less information-sharing. Similarly, power-priors allow a range of values to be used for α which determines the extent of information-sharing.

Whilst we believe that the above identification of ‘core’ relationships is exhaustive, the use of citation-mining techniques may have missed relevant methods, particularly those outside of health research. Additionally, we only looked for methods that shared information between sources of evidence that address different research questions. Hence, methods such as commensurate priors which have been used to combine individual-patient data and aggregate-level evidence on the same research question [99] could also be useful for combining evidence sources that pertain to different research questions, but were here considered outside of the scope of the search.

This paper is the first to summarise and categorise the existing literature by classifying methods according to the ‘core’ assumption that they use to facilitate information-sharing. Further research could explore the following questions: first, how can we determine whether indirect evidence is relevant? Second, how can the appropriateness of each information-sharing method be assessed for the synthesis problem at hand? Finally, can the extent of information-sharing be quantified to assist transparent decision-making?

Conclusions

We conclude that a plethora of methods has been used to facilitate information-sharing. These can be categorised according to the main assumption they impose into functional, exchangeability-based, prior-based, and multivariate relationships. Despite the wide range of available methods, these are often used preferentially without ensuring that all options have been explored. Given that methods may differ in the degree of information-sharing they impose, the implication is that the chosen method may impose stronger or weaker information-sharing that what is considered appropriate by policy-makers. Further research should investigate ways of judging the appropriateness of the degree

of information-sharing imposed by each method, and assess the impact of using different methods on decisions.

Abbreviations

DAM Decision Analytic Model.
EMA European Medicines Agency.
FDA U.S. Food and Drugs Administration.
HTA Health Technology Assessment.
MA Meta-Analysis.
NHS National Health Service.
NICE National Institute for Health and Care Excellence.
NMA Network Meta-Analysis.
QoL Quality of Life.
RE Random-Effects.
RTE Relative Treatment Effect.
TA Technology Appraisal.

Declarations

Ethics approval and consent to participate
 Not applicable.

Consent for publication

All authors have provided their consent for the publication of the final manuscript.

Availability of data and materials

The list of included studies in the systematic review is available online in the following url:

https://www.zotero.org/groups/2360368/citation-mining_included-studies

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by a doctoral studentship awarded to GFN by the Centre for Health Economics. The Centre for Health Economics did not have a role in the design of the study, the collection, analysis, and interpretation of data, or in writing the manuscript.

Authors' contributions

GFN drafted the initial manuscript, conducted the citation-mining review and coordinated contributions from all authors in drafting the final manuscript. MS oversaw the work. MS, BW, and SP participated in the study-design, revised the manuscript and approved its final version.

Acknowledgements

The authors acknowledge feedback received on related material during the first author's Thesis Advisory Panel meetings from Dr. Sylwia Bujkiewicz and Dr. Mona Kanaan.

References

1. World Health Organization: WHO HTA Definition (EB 134/30) (2018). <http://www.who.int/health-technology-assessment/about/Defining/en/>
2. Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., Devlin, N., Smith, P.C., Sculpher, M.: Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technology Assessment (Winchester, England)* **19**(14), 1–503 (2015)
3. Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L., Torrance, G.W.: *Methods for the Economic Evaluation of Health Care Programmes*, 4th edn. Oxford: Oxford University Press, ??? (2015)
4. Briggs, A., Claxton, K., Sculpher, M.J.: *Decision Modelling for Health Economic Evaluation*. Oxford University Press, ??? (2006)
5. Centre for Reviews and Dissemination: *Systematic reviews. crd's guidance for undertaking reviews in health care*. Technical report, Centre for Reviews and Dissemination (2006)

6. DerSimonian, R., Laird, N.: Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**(3), 177–188 (1986)
7. Lu, G., Ades, A.E.: Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* **23**(20), 3105–3124 (2004)
8. Ades, A.E., Sutton, A.J.: Multiparameter evidence synthesis in epidemiology and medical decision-making: current approaches. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **169**(1), 5–35 (2006)
9. Sweeting, M.J., Sutton, A.J., Lambert, P.C.: What to add to nothing? use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine* **23**(9), 1351–1375 (2004)
10. Int'Hout, J., Ioannidis, J.P.A., Borm, G.F., Goeman, J.J.: Small studies are more heterogeneous than large ones: a meta-meta-analysis. *Journal of Clinical Epidemiology* **68**(8), 860–869 (2015)
11. Food and Drug Administration, Center for Devices and Radiological Health: Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices (2016)
12. European Medicines Agency: Extrapolation of efficacy and safety in paediatric medicine development - EMA/199678/2016 (2016)
13. Ohlssen, D., Price, K.L., Amy Xia, H., Hong, H., Kerman, J., Fu, H., Quartey, G., Heilmann, C.R., Ma, H., Carlin, B.P.: Guidance on the implementation and reporting of a drug safety bayesian network meta-analysis. *Pharmaceutical Statistics* **13**(1), 55–70 (2014)
14. Duarte, A., Mebrahtu, T., Goncalves, P.: Adalimumab, etanercept and ustekinumab for treating plaque psoriasis in children and young people: systematic review and economic evaluation. *Health Technology Assessment* (2017)
15. Faria, R., Woods, B., Griffin, S., Palmer, S., Sculpher, M., Ryder, S.D.: Prevention of progression to cirrhosis in hepatitis c with fibrosis: effectiveness and cost effectiveness of sequential therapy with new direct-acting anti-virals. *Alimentary Pharmacology & Therapeutics* **44**(8), 866–876 (2016)
16. Corbett, M., Soares, M., Jhuti, G., Rice, S., Spackman, E., Sideris, E., Moe-Byrne, T., Fox, T., Marzo-Ortega, H., Kay, L., Woolcott, N., Palmer, S.: Tumour necrosis factor- α inhibitors for ankylosing spondylitis and non-radiographic axial arthritis: a systematic review and economic evaluation. *Health Technology Assessment* **20**(9) (2016)
17. McDaid, C., Griffin, S., Weatherly, H., Duree, K., van der Burgt, M.: Continuous positive airway pressure devices for the treatment of obstructive sleep apnoea-hypopnoea syndrome: a systematic review and economic analysis. *Health Technology Assessment* **13**(4) (2009)
18. Burch, J., Paulden, M., Conti, S., Stock, C., Corbett, M., Welton, N.J., Ades, A., Sutton, A., Cooper, N., Elliot, A., Nicholson, K., Duffy, S., McKenna, C., Stewart, L., Westwood, M., Palmer, S.: Antiviral drugs for the treatment of influenza: A systematic review and economic evaluation. *Health Technology Assessment* (2008)
19. Riemsma, R., Lhachimi, S.K., Armstrong, N., van Asselt, A.D.I., Allen, A., Manning, N., Harker, J., Tushabe, D.A., Severens, J.L., Kleijnen, J.: Roflumilast for the management of severe chronic obstructive pulmonary disease: A single technology appraisal. National Academies Press (US) (2011)
20. Higgins, J., Green, S.: *Cochrane handbook for systematic reviews of interventions* (2011)
21. Grandage, K., Slawson, D., Shaughnessy, A.F.: Site-ation pearl growing: methods and librarianship history and theory. *J Med Libr Assoc* **3**, 298–304 (2002)
22. Badampudi, D., Wohlin, C., Petersen, K.: Experiences from using snowballing and database searches in systematic literature studies. In: *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering. EASE '15*, pp. 17–11710. ACM, New York, NY, USA (2015)
23. Verde, P.E., Ohmann, C.: Combining randomized and non-randomized evidence in clinical research: a review of methods and applications. *Research Synthesis Methods* **6**(1), 45–62 (2015)
24. Higgins, J.P.T., Whitehead, A.: Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* **15**(24), 2733–2749 (1996)
25. Ades, A.E., Sculpher, M., Sutton, A., Abrams, K., Cooper, N., Welton, N., Lu, G.: Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics* **24**, 1–19 (2006)
26. Jackson, D., Riley, R., White, I.R.: Multivariate meta-analysis: Potential and promise. *Statistics in Medicine* **30**(20), 2481–2498 (2011)
27. Efthimiou, O., Debray, T.P.A., van Valkenhoef, G., Trelle, S., Panayidou, K., Moons, K.G.M., Reitsma, J.B., Shang, A., Salanti, G., on behalf of GetReal Methods Review Group: Getreal in network meta-analysis: a review of the methodology. *Research Synthesis Methods* **7**(3), 236–263 (2016)
28. Hobbs, B.P., Carlin, B.P., Mandrekar, S.J., Sargent, D.J.: Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* **67**(3), 1047–56 (2011)
29. Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., Neuenschwander, B.: Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics* **70**(4), 1023–1032 (2014)
30. Madan, J., Chen, Y.-F., Aveyard, P., Wang, D., Yahaya, I., Munafo, M., Bauld, L., Welton, N.: Synthesis of evidence on heterogeneous interventions with multiple outcomes recorded over multiple follow-up times reported inconsistently: a smoking cessation case-study. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **177**(1), 295–314 (2014)
31. Welton, N.J., Willis, S.R., Ades, A.E.: Synthesis of survival and disease progression outcomes for health technology assessment of cancer therapies. *Research Synthesis Methods* **1**(3–4), 239–257 (2010)
32. Dakin, H.A., Welton, N.J., Ades, A.E., Collins, S., Orme, M., Kelly, S.: Mixed treatment comparison of repeated measurements of a continuous endpoint: an example using topical treatments for primary open-angle glaucoma and ocular hypertension. *Stat Med* **30**(20), 2511–35 (2011)
33. Roever, C., Wandel, S., Friede, T.: Model averaging for robust extrapolation in evidence synthesis. *Statistics In Medicine* **38**(4, SI), 674–694 (2019)
34. Gamalo-Siebers, M., Savic, J., Basu, C., Zhao, X., Gopalakrishnan, M., Gao, A., Song, G., Baygani, S., Thompson, L., Xia, H.A., Price, K., Tiwari, R., Carlin, B.P.: Statistical modeling for Bayesian extrapolation of adult clinical trial information in pediatric drug evaluation. *Pharmaceutical Statistics* **16**(4), 232–249 (2017)
35. Soares, M.O., Dumville, J.C., Ades, A.E., Welton, N.J.: Treatment comparisons for decision making: facing the problems of sparse and few data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **177**(1), 259–279 (2014)
36. Achana, F.A., Cooper, N.J., Dias, S., Lu, G., Rice, S.J., Kendrick, D., Sutton, A.J.: Extending methods for investigating the relationship between treatment effect and baseline risk from pairwise meta-analysis to network meta-analysis. *Stat Med* **32**(5), 752–71 (2013)
37. Dias, S., Welton, N., Sutton, A., Ades, A.: NICE DSU Technical Support Document 5: Evidence Synthesis in the Baseline Natural History Model. NICE DSU Technical Support Document in Evidence Synthesis, vol. TSD5. National Institute for Health and Clinical Excellence, ??? (2011)
38. Mawdsley, D., Bennetts, M., Dias, S., Boucher, M., Welton, N.J.: Model-based network meta-analysis: A framework for evidence synthesis of clinical trial data. *CPT Pharmacometrics Syst Pharmacol* **5**(8), 393–401 (2016)
39. Del Giovane, C., Vacchi, L., Mavridis, D., Filippini, G., Salanti, G.: Network meta-analysis models to account for variability in treatment definitions: application to dose effects. *Stat Med* **32**(1), 25–39 (2013)
40. Owen, R.K., Tincello, D.G., Keith, R.A.: Network meta-analysis: development of a three-level hierarchical modeling approach incorporating dose-related constraints. *Value Health* **18**(1), 116–26 (2015)
41. Wu, J., Banerjee, A., Jin, B., Menon, S.M., Martin, S.W., Heatherington, A.C.: Clinical dose-response for a broad set of biological products: A model-based meta-analysis. *Statistical Methods in Medical Research* **27**(9), 2694–2721 (2018)
42. Langford, O., Aronson, J.K., van Valkenhoef, G., Stevens, R.J.: Methods for meta-analysis of pharmacodynamic dose-response data

- with application to multi-arm studies of alogliptin. *Statistical Methods in Medical Research* **27**(2), 564–578 (2018)
43. da Costa, B.R., Reichenbach, S., Keller, N., Nartey, L., Wandel, S., Juni, P., Trelle, S.: Effectiveness of non-steroidal anti-inflammatory drugs for the treatment of pain in knee and hip osteoarthritis: a network meta-analysis. *Lancet* **390**(10090), 21–33 (2017)
 44. Warren, F.C., Abrams, K.R., Sutton, A.J.: Hierarchical network meta-analysis models to address sparsity of events and differing treatment classifications with regard to adverse outcomes. *Statistics In Medicine* **33**(14), 2449–2466 (2014)
 45. Nixon, R.M., Bansback, N., Brennan, A.: Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Stat Med* **26**(6), 1237–54 (2007)
 46. Dominici, F., Parmigiani, G., Wolpert, R.L., Hasselblad, V.: Meta-analysis of migraine headache treatments: Combining information from heterogeneous designs. *Journal of the American Statistical Association* **94**(445), 16–28 (1999)
 47. Moreno, S.G., Sutton, A.J., Ades, A.E., Cooper, N.J., Abrams, K.R.: Adjusting for publication biases across similar interventions performed well when compared with gold standard data. *Journal of Clinical Epidemiology* **64**(11), 1230–1241 (2011)
 48. Welton, N.J., Caldwell, D.M., Adamopoulos, E., Vedhara, K.: Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. *Am J Epidemiol* **169**(9), 1158–65 (2009)
 49. Mills, E.J., Thorlund, K., Ioannidis, J.P.A.: Calculating additive treatment effects from multiple randomized trials provides useful estimates of combination therapies. *Journal of Clinical Epidemiology* **65**(12), 1282–1288 (2012)
 50. Melendez-Torres, G.J., Bonell, C., Thomas, J.: Emergent approaches to the meta-analysis of multiple heterogeneous complex interventions. *BMC Medical Research Methodology* **15** (2015)
 51. Efthimiou, O., Mavridis, D., Debray, T.P.A., Samara, M., Belger, M., Siontis, G.C.M., Leucht, S., Salanti, G., on behalf of GetReal Work, P.: Combining randomized and non-randomized evidence in network meta-analysis. *Statistics in Medicine* **36**(8), 1210–1226 (2017)
 52. Schmitz, S., Adams, R., Walsh, C.: Incorporating data from various trial designs into a mixed treatment comparison model. *Statistics in Medicine* **32**(17), 2935–2949 (2013)
 53. Rietbergen, C.: Quantitative evidence synthesis with power priors. PhD thesis, University of Utrecht (2016)
 54. Prevost, T., Abrams, K., Jones, D.: Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. *Statistics In Medicine* **19**(24), 3359–3376 (2000)
 55. Turner, R.M., Spiegelhalter, D.J., Smith, G.C.S., Thompson, S.G.: Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **172**(1), 21–47 (2009)
 56. Spiegelhalter, D.J., Best, N.G.: Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Stat Med* **22**(23), 3687–709 (2003)
 57. Dias, S., Welton, N.J., Marinho, V.C.C., Salanti, G., Higgins, J.P.T., Ades, A.E.: Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **173**(3), 613–629 (2010)
 58. Welton, N.J., Ades, A.E., Carlin, J.B., Altman, D.G., Sterne, J.A.C.: Models for Potentially Biased Evidence in Meta-Analysis Using Empirically Based Priors. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **172**(1), 119–136 (2009)
 59. Chaimani, A., Salanti, G.: Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. *Research Synthesis Methods* **3**(2), 161–176 (2012)
 60. Trinquart, L., Chatellier, G., Ravaud, P.: Adjustment for reporting bias in network meta-analysis of antidepressant trials. *BMC Medical Research Methodology* **12**, 150–150 (2012)
 61. Mavridis, D., Sutton, A., Cipriani, A., Salanti, G.: A fully bayesian application of the copas selection model for publication bias extended to network meta-analysis. *Stat Med* **32**(1), 51–66 (2013)
 62. Salanti, G., Dias, S., Welton, N.J., Ades, A.E., Gollinopoulos, V., Kyrgiou, M., Mauri, D., Ioannidis, J.P.: Evaluating novel agent effects in multiple-treatments meta-regression. *Stat Med* **29**(23), 2369–83 (2010)
 63. Eddy, D.M., Hasselblad, V., Shachter, R.: An introduction to a bayesian method for meta-analysis: The confidence profile method. *Med Decis Making* **10**(1), 15–23 (1990)
 64. Wolpert, R.L., Kerrie, L.M.: Adjusted likelihoods for synthesizing empirical evidence from studies that differ in quality and design: Effects of environmental tobacco smoke. *Statistical Science* **19**(3), 450–471 (2004)
 65. Welton, N.J., Cooper, N.J., Ades, A.E., Lu, G., Sutton, A.J.: Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza a and b. *Stat Med* **27**(27), 5620–39 (2008)
 66. Lu, G., Ades, A.E., Sutton, A.J., Cooper, N.J., Briggs, A.H., Caldwell, D.M.: Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Stat Med* **26**(20), 3681–99 (2007)
 67. Ding, Y., Fu, H.: Bayesian indirect and mixed treatment comparisons across longitudinal time points. *Stat Med* **32**(15), 2613–28 (2013)
 68. Jackson, D., Rollins, K., Coughlin, P.: A multivariate model for the meta-analysis of study level survival data at multiple times. *Research Synthesis Methods* **5**(3), 264–272 (2014)
 69. Musekiwa, A., Manda, S.O.M., Mwambi, H.G., Chen, D.-G.: Meta-Analysis of Effect Sizes Reported at Multiple Time Points Using General Linear Mixed Model. *PLOS ONE* **11**(10) (2016)
 70. Bujkiewicz, S., Thompson, J.R., Riley, R.D., Abrams, K.R.: Bayesian meta analytical methods to incorporate multiple surrogate endpoints in drug development process. Hoboken (2015)
 71. Bujkiewicz, S., Thompson, J.R., Sutton, A.J., Cooper, N.J., Harrison, M.J., Symmons, D.P.M., Abrams, K.R.: Use of bayesian multivariate meta-analysis to estimate the haq for mapping onto the eq-5d questionnaire in rheumatoid arthritis. *Value in Health* (2014)
 72. Mavridis, D., Salanti, G.: A practical introduction to multivariate meta-analysis. *Stat Methods Med Res* **22**(2), 133–58 (2013)
 73. van Houwelingen, H., Zwinderman, K., Stijnen, T.: A bivariate approach to meta-analysis. *Statistics In Medicine* **12**, 2273–2284 (1993)
 74. van Houwelingen, H., Arends, L., Stijnen, T.: Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics In Medicine* **21**(4), 589–624 (2002)
 75. Efthimiou, O., Mavridis, D., Cipriani, A., Leucht, S., Bagos, P., Salanti, G.: An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. *Statistics in Medicine* **33**(13), 2275–2287 (2014)
 76. Ades, A.E., Mavranzouli, I., Dias, S., Welton, N.J., Whittington, C., Kendall, T.: Network meta-analysis with competing risk outcomes. *Value Health* **13**(8), 976–83 (2010)
 77. Achana, F.A., Cooper, N.J., Bujkiewicz, S., Hubbard, S.J., Kendrick, D., Jones, D.R., Sutton, A.J.: Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. *BMC Med Res Methodol* **14**, 92 (2014)
 78. Hong, H., Chu, H., Zhang, J., Carlin, B.P.: A bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Research Synthesis Methods* **7**(1), 6–22 (2016)
 79. Riley, R.D., Thompson, J.R., Abrams, K.R.: An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics* **9**(1), 172–86 (2008)
 80. Daniels, M.J., Hughes, M.D.: Meta-analysis for the evaluation of potential surrogate markers. *Stat Med* **16**(17), 1965–82 (1997)
 81. Jackson, D., Bujkiewicz, S., Law, M., Riley, R.D., White, I.R.: A matrix-based method of moments for fitting multivariate network meta-analysis models with multiple outcomes and random inconsistency effects. *Biometrics* **74**(2), 548–556 (2018)
 82. Jackson, D., White, I.R., Riley, R.D.: A matrix-based method of moments for fitting the multivariate random effects model for meta-analysis and meta-regression. *BIOMETRICAL JOURNAL* **55**(2), 231–245 (2013)
 83. Jackson, D., Riley, R.D.: A refined method for multivariate meta-analysis and meta-regression. *Statistics In Medicine* **33**(4), 541–554 (2014)
 84. Nam, I., Mengersen, K., Garthwaite, P.: Multivariate meta-analysis. *Statistics In Medicine* **22**(14), 2309–2333 (2003)

85. Riley, R.D., Abrams, K.R., Lambert, P.C., Sutton, A.J., Thompson, J.R.: An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics In Medicine* **26**(1), 78–97 (2007)
86. Wei, Y., Higgins, J.P.T.: Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Statistics In Medicine* **32**(7), 1191–1205 (2013)
87. Wei, Y., Higgins, J.P.T.: Bayesian multivariate meta-analysis with multiple outcomes. *Statistics In Medicine* **32**(17), 2911–2934 (2013)
88. Turner, R.M., Jackson, D., Wei, Y., Thompson, S.G., Higgins, J.P.T.: Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics In Medicine* **34**(6), 984–998 (2015)
89. Pullenayegum, E.M.: An informed reference prior for between-study heterogeneity in meta-analyses of binary outcomes. *Statistics In Medicine* **30**(26), 3082–3094 (2011)
90. Rhodes, K.M., Turner, R.M., Higgins, J.P.T.: Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *Journal of Clinical Epidemiology* **68**(1), 52–60 (2015)
91. Cooper, N.J., Sutton, A.J., Morris, D., Ades, A.E., Welton, N.J.: Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Statistics in Medicine* **28**(14), 1861–1881 (2009)
92. Dias, S., Welton, N., Sutton, A., Ades, A.: NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. Technical Support Document in Evidence Synthesis, vol. TSD2. National Institute for Health and Clinical Excellence, ??? (2011)
93. Dias, S., Sutton, A., Welton, N., Ades, A.: NICE DSU Technical Support Document 3: Heterogeneity: Subgroups, Meta-Regression, Bias and Bias-Adjustment. National Institute for Health and Clinical Excellence, ??? (2011)
94. Lu, G., Ades, A.: Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* **10**(4), 792–805 (2009)
95. Thorlund, K., Thabane, L., Mills, E.J.: Modelling heterogeneity variances in multiple treatment comparison meta-analysis. are informative priors the better solution? *BMC Medical Research Methodology* **13**, 2–2 (2013)
96. Bujkiewicz, S., Thompson, J.R., Riley, R.D., Abrams, K.R.: Bayesian meta-analytical methods to incorporate multiple surrogate endpoints in drug development process. *Statistics In Medicine* **35**(7, SI), 1063–1089 (2016)
97. Copas, J.B., Jackson, D., White, I.R., Riley, R.D.: The role of secondary outcomes in multivariate meta-analysis. *Journal of the Royal Statistical Society Series C - Applied Statistics* **67**(5), 1177–1205 (2018)
98. Efthimiou, O., Mavridis, D., Riley, R.D., Cipriani, A., Salanti, G.: Joint synthesis of multiple correlated outcomes in networks of interventions. *BIostatistics* **16**(1), 84–97 (2015)
99. Hong, H., Fu, H., Carlin, B.P.: Power and commensurate priors for synthesizing aggregate and individual patient level data in network meta-analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* (2018)
100. Hwang, H., DeSantis, S.M.: Multivariate network meta-analysis to mitigate the effects of outcome reporting bias. *Statistics In Medicine* **37**, 3254–3266 (2018)
101. Liu, Y., DeSantis, S.M., Chen, Y.: Bayesian mixed treatment comparisons meta-analysis for correlated outcomes subject to reporting bias. *Journal of the Royal Statistical Society Series C - Applied Statistics* **67**(1), 127–144 (2018)
102. Lu, G., Kounali, D., Ades, A.E.: Simultaneous Multioutcome Synthesis and Mapping of Treatment Effects to a Common Scale. *VALUE IN HEALTH* **17**(2), 280–287 (2014)
103. Tan, S.H., Abrams, K.R., Bujkiewicz, S.: Bayesian Multiparameter Evidence Synthesis to Inform Decision Making: A Case Study in Metastatic Hormone-Refractory Prostate Cancer. *Medical Decision Making* **38**(7), 834–848 (2018)
104. Salanti, G., Marinho, V., Higgins, J.P.: A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol* **62**(8), 857–64 (2009)
105. McCarron, C.E., Pullenayegum, E.M., Thabane, L., Goeree, R., Tarride, J.-E.: The importance of adjusting for potential confounders in Bayesian hierarchical models synthesising evidence from randomised and non-randomised studies: an application comparing treatments for abdominal aortic aneurysms. *BMC Medical Research Methodology* **10** (2010)
106. Mak, A., Cheung, M.W.L., Ho, R.C.-M., Cheak, A.A.-C., Lau, C.S.: Bisphosphonates and atrial fibrillation: Bayesian meta-analyses of randomized controlled trials and observational studies. *BMC Musculoskeletal Disorders* **10** (2009)
107. McCarron, C.E., Pullenayegum, E.M., Thabane, L., Goeree, R., Tarride, J.-E.: Bayesian Hierarchical Models Combining Different Study Types and Adjusting for Covariate Imbalances: A Simulation Study to Assess Model Performance. *PLOS ONE* **6**(10) (2011)
108. Ren, S., Oakley, J.E., Stevens, J.W.: Incorporating Genuine Prior Information about Between-Study Heterogeneity in Random Effects Pairwise and Network Meta-analyses. *Medical Decision Making* **38**(4), 531–542 (2018)
109. Ibrahim, J.G., Chen, M.-H.: Power prior distributions for regression models. *Statist. Sci.* (2000)
110. Rietbergen, C., Groenwold, R.H.H., Hoijtink, H.J.A., Moons, K.G.M., Klugkist, I.: Expert elicitation of study weights for bayesian analysis and meta-analysis. *Journal of Mixed Methods Research* **10**(2), 168–181 (2016)
111. Spiegelhalter, J.D., Abrams, R.K., Myles, P.J.: *Bayesian Approaches to Clinical Trials and Health-care Evaluation*. Wiley, ??? (2004)
112. Hong, C., Riley, R.D., Chen, Y.: An improved method for bivariate meta-analysis when within-study correlations are unknown. *Research Synthesis Methods* **9**(1), 73–88 (2018)
113. Bujkiewicz, S., Thompson, J.R., Sutton, A.J., Cooper, N.J., Harrison, M.J., Symmons, D.P.M., Abrams, K.R.: Multivariate meta-analysis of mixed outcomes: a bayesian approach. *Statistics in Medicine* **32**(22), 3926–3943 (2013)
114. Jackson, D., White, I.R., Price, M., Copas, J., Riley, R.D.: Borrowing of strength and study weights in multivariate and network meta-analysis. *Statistical Methods in Medical Research* **26**(6), 2853–2868 (2017)

Figure legends

Figure 1 An illustration of the extended evidence base The small pie in the middle, characterised by characterised by P^0, I^0, C^0, O^0, S^0 , represents only the directly relevant information which usually comprise only a small part of the evidence that is relevant to a decision. The evidence outside the small pie represent examples of indirectly relevant information for each of the PICOS levels.

Figure 2 PRISMA diagram for search results

Figure 3 'Core' categories of information-sharing

Figure 4 An illustration of a multi-level model

Table legends

Table 1 A categorisation of papers that share information on the relative effectiveness parameter according to the 'core' relationship that they use and the PICOS level that direct and indirect evidence differ in

Additional Files

Additional file 1 — Search strategy.

A description of the inclusions and exclusion criteria of the search as well as the number of citations for each one of the seminal papers/‘pearls’ and the number of times it has been cited.

Additional file 2 — PRISMA checklist for systematic reviews.

The PRISMA checklist for systematic reviews, indicating the page at which each characteristic of the review is described.

Additional file 3 — Brief summary of each included study.

A multi-page table including a brief description of how each of the included papers shared information between direct and indirect information and on which PICOS level.