Clarifying Values: An Updated and Expanded Systematic Review and Meta-Analysis

Holly O. Witteman1,2,3, Ruth Ndjaboue1,2, Gratianne Vaisson1, Selma Chipenda Dansokho1, Bob Arnold4, John F. P. Bridges5, Sandrine Comeau1, Angela Fagerlin6, Teresa Gavaruzzi7, Melina Marcoux1, Arwen Pieterse8, Michael Pignone9, Thierry Provencher1, Charles Racine1, Dean Regier10, Charlotte Rochefort-Brihay1, Praveen Thokala11, Marieke Weernink12, Douglas B. White13, Celia E. Wills14, Jesse Jansen15

1Department of Family and Emergency Medicine, Facul ty of Medicine, Laval University, Quebec City, Canada
2VITAM Research Centre, Quebec City, Canada
3CHU de Québec Research Centre, Quebec City, Canada
4UPMC Palliative and Supportive Institute; University of Pittsburgh, Division of General Internal Medicine, Section of Palliative care and Medical ethics, Pittsburgh, PA, USA
5Department of Biomedical Informatics, The Ohio State University College of Medicine, Columbus, OH, USA
6Department of Population Health Sciences, University of Utah School of Medicine, Salt Lake City, UT, USA
7Department of Developmental Psychology and Socialization, University of Padova, Padova, Italy
8Leiden University Medical Center, Leiden, The Netherlands
9Departments of Internal Medicine and Population Health, Dell Medical School, University of Texas, Austin, TX, USA
10School of Population and Public Health, Univ. of British Columbia, Vancouver, BC, Canada
11School of Health and Related Research, University of Sheffield, Sheffield, UK
12Municipal Health Services (GGD), Enschede, The Netherlands.
13Program on Ethics and Decision Making in Critical Illness, Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
14College of Nursing, Center on Healthy Aging, Self-Management and Complex Care, Ohio State University, Columbus, OH, USA
15Department of Family Medicine/CAPHRI, Maastricht University, Maastricht, The Netherlands

Corresponding author: Holly O. Witteman, 1050 avenue de la Médecine, Pavillon Ferdinand-Vandry, Université Laval, Québec City, Québec, G1V 0A6, Canada, telephone: 418.656.2131 x3981, fax: 418-656.2465, email: holly.witteman@fmed.ulaval.ca

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Background: Patient decision aids should help people make evidence-informed decisions aligned with their values. There is limited guidance about how to achieve such alignment.

Purpose: To describe the range of values clarification methods available to patient decision aid developers, synthesize evidence regarding their relative merits, and foster collection of evidence by offering researchers a proposed set of outcomes to report when evaluating the effects of values clarification methods.

Data Sources: MEDLINE, EMBASE, PubMed, Web of Science, the Cochrane Library, CINAHL

Study Selection: We included articles that described randomized trials of one or more explicit values clarification methods. From 30,648 records screened, we identified 33 articles describing trials of 43 values clarification methods.

Data Extraction: Two independent reviewers extracted details about each values clarification method and its evaluation.

Data Synthesis: Compared to control conditions or to implicit values clarification methods, explicit values clarification methods decreased the frequency of values-disgruent choices (risk difference -0.04 95% CI [-0.06 to -0.02], p<.001) and decisional regret (standardized mean difference -0.20 95% CI [-0.29 to -0.11], p<0.001). Multicriteria decision analysis led to more values-congruent decisions than other values clarification methods (Chi-squared(2)=9.25, p=.01). There were no differences between different values clarification methods regarding decisional conflict (Chi-squared(2)=6.08, p=.05).

Limitations: Some meta-analyses had high heterogeneity. We grouped values clarification methods into broad categories.

Conclusions: Current evidence suggests patient decision aids should include an explicit values clarification method. Developers may wish to specifically consider multicriteria decision analysis. Future evaluations of values clarification methods should report their effects on decisional conflict, decisions made, values congruence, and decisional regret.
Introduction

Shared decision making aims to foster health-related decisions that are both informed by the best available evidence and aligned with what matters to the person or people affected by the decision. [1–4] Individual values are a critical ingredient in high quality individual health decision making. [5–7] What is important to one person might be different from what is important to others, and determining what is important to oneself can be difficult even if one has the appropriate information and evidence at hand. Therefore, patient decision aids should both present evidence appropriately and also support the process of clarifying and expressing patients’ (and, when appropriate, other relevant stakeholders’) values, with the goal of supporting alignment between values and decisions. Within patient decision aids, such support is offered by explicit values clarification methods.

Explicit values clarification methods require users to interact with something such as a worksheet or an interactive website to clarify what matters to them relevant to a health decision. Such methods have been shown to encourage desirable outcomes such as better alignment with patients’ values [8, 9] and reduced decisional regret, the latter particularly among people with lower health literacy. [10] However, explicit values clarification methods are extremely diverse, [11], and there has been little guidance regarding their comparative effects on users’ decision-making processes or outcomes [12], making it difficult for patient decision aid developers to know which explicit method to use. Patient decision aid developers might look towards the preference elicitation literature for guidance, but the guidance available [13] is often tailored towards aggregate level decision making, such as regulatory decisions [14] or health technology assessment [15], not for supporting individual-level decision making.

This updated review sought to build upon previous versions of the International Patient Decision Aids Standards’ Chapter on Values Clarification [16, 17] as well as previous evidence syntheses that have established the advantages of explicit values clarification methods over implicit methods or no values clarification. [8, 9] We sought to advance the science and practice of values clarification methods in three ways. First, we aimed to offer clear definitions and an annotated summary of existing approaches that have been or could be used as values clarification methods. Second, we aimed to synthesize evidence of different techniques’ effects on health decision outcomes. Third, we aimed to foster future evidence by offering researchers a proposed set of outcomes to consider when evaluating the effects of values clarification methods.

Definitions

Part of the challenge in studying or using values clarification methods is that definitions vary and terms like ‘values’ are used imprecisely in the patient decision support literature. [18, 19] Another challenge is that there is substantial overlap between values clarification methods used in patient decision support and preference elicitation methods used in health economics. To bring clarity to this imprecision and overlap, we adopt working definitions in Table 1 for use in this paper.

Table 1. Definitions of Terms

| Term | Definition adopted in this paper |
|------|---------------------------------|
| Values | An umbrella term referring to what matters to an individual relevant to a health decision. Values may be directly relevant to decisions (e.g., “beliefs, feelings, or perceptions regarding attributes of a treatment option”) or indirectly relevant (e.g., goals; worldviews; family, religious, or cultural values). [20] Values may be represented qualitatively or, in some cases, quantitatively. This definition is |
| Values clarification methods | “Strategies that are intended to help patients evaluate the desirability of options or attributes of options within a specific decision context, in order to identify which option [they] prefer.” [17] |
|-----------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| Implicit values clarification methods | Strategies for facilitating values clarification that do not require people to interact with anything or anyone; e.g., describing “options in enough detail that clients can imagine what it is like to experience the physical, emotional, and social effects,” [9] or simply encouraging people to think about what matters to them. |
| Explicit values clarification methods | Strategies for facilitating values clarification that require people to interact with something or someone; e.g., filling out a worksheet, using an interactive website, having a semi-structured conversation with another person with the explicit purpose of clarifying values, or engaging in another structured exercise. |
| Preferences | The extent to which a decision option or health state is desirable or acceptable, either in the abstract or in comparison to other options or health states. Preferences may be represented qualitatively or, more commonly, quantitatively. [21] |
| Preference elicitation methods | Processes by which preferences are drawn out. [11] Preference elicitation methods may vary according to the theory informing them. They are highly related to values clarification methods. Although older terms “revealed” and “stated” preference elicitation methods are no longer recommended, readers who encounter these terms in previous preference elicitation literature should note that these may overlap with implicit and explicit values clarification methods, respectively. |

As noted above, we continue to use the term values clarification even though this is sometimes misinterpreted as implying a narrow definition of values. Changing terms makes it difficult for people who are new to a field to connect the dots across decades of previous research. It is clear that previous research in values clarification addressed issues that were broader than valuation of treatment-specific attributes. [16] In this update, we therefore move forward with the older terms, now with more clarity about what they mean in our presentation of the evidence.

**Theoretical Rationale**

Our interdisciplinary team determined that the theoretical rationale for values clarification required only a small edit, shown in square brackets, to reflect the focus on explicit methods. Like Fagerlin and colleagues, we assert the theoretical rationale for explicit values clarification methods as being that they, “should aim to [explicitly] facilitate at least one or more of the following six decision-making processes: 1) Identifying options, which can include either the narrowing down of options, or the generation of options that were not offered at the outset, 2) Identifying attributes of the situation and/or the options which ultimately affect the patient’s preference in a specific decision context, 3) Reasoning about options or attributes of options, 4) Integrating attributes of options using either compensatory or both compensatory and non-compensatory decision rules, 5) Making holistic comparisons, and 6) Helping decision makers retrieve relevant values from long-term memory.” [17] Pieterse and colleagues provided theory-based recommendations on processes that values clarification methods could aim to facilitate. [22]
Although reasoning is one of the potential processes supported by values clarification, neither the definition nor the theoretical rationale of values clarification methods requires that people who are being supported in making a personal health decision must rationally deliberate about each option, nor that the goal must always be a fully rational choice. In some decision-making situations, rational deliberation and rational choice may be desired, while in others, they may not. [23, 24]

Explicit Values Clarification Methods

Table 2 organizes strategies that can be used as explicit values clarification methods in patient decision aids, building upon previously-developed lists of types of values clarification methods [7, 11] and reviews of preference elicitation methods. [25, 26] Methods range from highly structured strategies that can also be used for preference elicitation in the context of health policy decision making to substantially less structured strategies. While not every use of a given method will be exactly the same, we deemed them functionally similar in terms of how they might be used and what the user experience might be in a patient decision aid. Patient decision aids may use multiple strategies. For example, a user may be asked to use a rating scale or visual analog scale whose values are then used in a decision analytic model.

Table 2. Values Clarification Methods

| Method                                      | Description                                                                 |
|---------------------------------------------|----------------------------------------------------------------------------|
| Adaptive Conjoint Analysis (example [27])   | The user rates a series of sets of attributes and their levels, where choices presented are tailored to earlier answers. |
| Allocation of Points (example [28])         | The user has a “budget” to “spend” on decision attributes, according to their importance. |
| Analytical Hierarchy Process (example [29]) | The user is asked to compare sets of options relative to predefined decision criteria. |
| Best-Worst Scaling (example [30])           | The user is asked to indicate the best and the worst in sets of options with different attributes and levels. |
| Decision Analysis or Multicriteria Decision Analysis (umbrella term*) (resource [31, 32]) | The user is asked to directly indicate the extent to which a decision attribute or outcome matters to them or how good or bad they deem it to be. These values are then used in a model that calculates alignment between what matters to the user and the available decision options. |
| Discrete Choice Experiments (example [33])  | The user is asked to make a series of choices between two (or more) alternatives, where each alternative is characterized by attributes and their associated levels. |
| Open Discussion (example [34])              | The user makes a list and/or discusses what matters to them in an unstructured or semi-structured discussion. |
| Pros and Cons (resource [35])               | The user lists advantages (pros) and disadvantages (cons) of options and/or indicates the relevance (‘this matters to me’) or importance (e.g., on a Likert scale) of each advantage or disadvantage. |
| Ranking (example [36])                      | The user is asked to place attributes in order of importance, relative to each other. |
| Rating Scales (example [37])                | The user indicates the importance of an attribute on a visual analog scale (e.g., paper-based visual analog scale, online slider) or Likert scale approximating a visual analog scale. If the rating is then used to calculate and show which option fits best, the method is classified as (multicriteria) decision analysis. |
Social Matching (example [38])
The user “observes different characters’ decisions and/or decision-making processes and identifies 1 or more characters” with whom they identify. [11]

Standard Gamble (example [39])
The user indicates their willingness to accept a certain risk of death in order to avoid a particular health state by choosing between the certainty of living in that health state for the remainder of their life versus a gamble between two possible outcomes: life in a state of optimal health, with probability p, or immediate death, with probability (1−p).

Time Tradeoff (example [39])
The user indicates how many years in their current health state they would be willing to ‘trade off’, in order to regain full health.

*Multicriteria decision analysis or decision analysis is an umbrella term. It encompasses some of the other, more specific categories (e.g., discrete choice experiments, best-worst scaling.) When applicable, we use the more specific, narrower categories. Otherwise, we use the umbrella term “multicriteria decision analysis,” or, for brevity in figures, “decision analysis.” Additionally, although within multicriteria decision analysis, the user may be asked to rate attributes on rating scales, what distinguishes multicriteria decision analysis from other methods such as rating scales is that the model calculates how well or poorly the options align with what matters to a user.

Methods

Our overall methods were guided by the Cochrane Handbook. We report according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [40] guidelines.

Eligibility criteria

We included published reports of comparative evaluations of explicit values clarification methods, whether they were called ‘values clarification methods’ in the publications or not. This meant that we included trials of preference elicitation methods that had been trialed as values clarification methods; for example, multi-criteria decision analysis or discrete choice experiments. We included evaluations using comparative methods; i.e., randomized controlled trials or randomized experiments of one or more values clarification methods. The comparisons could be one or more values clarification methods compared to a control method, or compared to each other. Because we sought to understand the effects of values clarification methods, we excluded evaluations using descriptive study designs (e.g., acceptability and feasibility study, development study), observational study designs (e.g., reporting outcomes before and after use of a values clarification method), and reports of values clarification methods that did not evaluate the method independently of the patient decision aid in which it was used. Randomized experiments comparing one or more values clarification methods had to use distinctly different methods, meaning that more than the content or presentation of information in the values clarification method varied.

We did not apply language restrictions. We applied date restrictions to the portion of the review for which we had already conducted a systematic review (i.e., evaluations of values clarification methods that used the term ‘values clarification’. [12, 17] Specifically, for this subgroup, we added articles indexed or published starting in 2014 to the existing set of articles indexed or published prior to 2014 that we had already identified using the same search strategy. We applied no date restrictions to the new, expanded portion of the review (i.e., evaluations of values clarification methods that did not use the term ‘values clarification’).

Information sources

We performed a systematic literature search in MEDLINE, EMBASE, Web of Science, the Cochrane Library, and CINAHL.
Search strategy

We developed a draft search strategy in collaboration with an information specialist (FB, see Acknowledgments). Search strategies for each database are shown in Online Appendix 1. We reviewed search strategies with all authors to ensure they were inclusive of relevant preference elicitation methods that might be used for values clarification. We conducted hand searches by reviewing articles that cited the previous version of these standards (values clarification chapter) or a previous systematic review of values clarification methods.

Study records: Data management

We managed data with Covidence (covidence.org, Melbourne, Australia), reviewing data records at regular team meetings.

Study records: Selection process

Two independent reviewers (SC, MM, TP, CR, CRB) screened titles and abstracts to assess potential relevance, with a third reviewer adjudicating discrepancies and discussions of questions and points of disagreement at regular team meetings. Two independent reviewers then reviewed the full text of all articles deemed potentially relevant based on their title and abstract. Discrepancies in inclusion and exclusion at full text were adjudicated through team discussions at regular meetings until we reached consensus.

Study records: Data collection process

Two independent, trained research team members (SC, MM, TP, CR, CRB) extracted data from each article using a standardized and pilot-tested data extraction form based on a previous form [12] and adapted to this review. We resolved lack of agreement through discussion until consensus was reached. We contacted authors to collect any needed data that they did not report or were unable to report in their publication.

Data items

Regarding study participants, we recorded the sample size for control and intervention groups along with basic inclusion and exclusion criteria and whether or not they were making the actual decision or if the study was hypothetical. We defined a hypothetical scenario as one in which people are asked (explicitly or implicitly) to imagine that they are in a certain situation or facing a certain decision. We defined a real scenario as one in which people are facing a decision (e.g., because they have received a diagnosis) or are members of a population likely to face the decision in the near term (e.g., parents of children eligible to receive vaccines within the coming months.)

Regarding interventions, we recorded the type of explicit values clarification method as listed in Table 2. We also recorded specific characteristics of each values clarification method, namely, whether it explicitly requires the user to engage with tradeoffs (i.e., considering which potential harms are acceptable in exchange for their associated potential benefits), whether it explicitly shows the user the correspondence between their options and what they value, and which, if any, theoretical or conceptual framework underpins it. Where relevant, we recorded whether a variable was collected via self-report, meaning whether responses were completed by participants themselves, or by independent researchers based on direct observation, including coded qualitative data.

For comparators (controls) we recorded whether the comparator was no values clarification method or an implicit method, and treated both as equivalent controls. The Cochrane review of patient decision aids...
specifies that all patient decision aids must contain implicit values clarification methods at minimum [9] and it is accordingly rare to have patient decision aids that do not present potential benefits and harms of options in organized ways. In other words, in the context of patient decision aids, there is no meaningful distinction between implicit methods and no values clarification. The different terminology is simply a function of how authors choose to name their control. We also recorded studies that compared different types of explicit values clarification methods to each other.

Outcomes

Whenever such data were available, we extracted data regarding values congruence as our primary outcome, as well as secondary outcomes: decision readiness (worry, decision uncertainty, decision-making preparation, knowledge); decisional conflict; decision made; post-decision and post implementation health and well-being (decisional regret, longer-term health outcomes). Following data extraction by pairs of trained reviewers (SC, MM, TP, CR, CRB), three authors (HOW, SCD, JJ) mapped all outcomes into broad outcome groups: worry (including perceived risk), decision uncertainty (not including decisional conflict), decisional conflict (decisional conflict scale or any subscales), decision-making preparation (including self-efficacy for decision-making), beliefs (including beliefs about the condition or underlying decision structure), knowledge, values (including reported utilities), shared decision making, effects on communication (including quality, length, or existence of communication), satisfaction with care, preferences, decision (choice made and implemented) or decisional intent (choice intended, or made and not yet implemented), values congruence, informed decision making, post-decision feelings (including satisfaction, regret), post-decision health, and user assessment of the intervention (including acceptability, satisfaction, perceived balance.) We conducted meta-analyses on primary outcome values congruence and secondary outcome decisional conflict, as these outcomes had sufficient studies to do so.

Risk of bias in individual studies

Independent, trained research team members assessed risk of bias for each study using methods as defined in the Cochrane Handbook, section 8.5. [41] We conducted quantitative data syntheses with and without studies identified as being at high risk of bias to determine the sensitivity of overall findings to these studies.

Data synthesis

We synthesized frequency-based results (e.g., how many values clarification methods reflect a given design) descriptively. To synthesize effects on outcomes, we pooled all experiments that evaluated a values clarification method against no values clarification method or an implicit method. For multi-armed studies in which the comparison of a decision aid with and without a values clarification method included an arm that was not relevant to our comparison of interest (for example, an information booklet serving as a control condition in an evaluation of the decision aid) we ignored the third arm. For multi-armed studies containing two or more different values clarification methods and one arm of implicit values clarification or control, we considered each comparison of a values clarification method against implicit values clarification, meaning that each of the multi-armed studies included in this review contributed multiple comparisons to the pooled set.

To meta-analyze results for values congruence, we pooled results using risk differences and applying a random effects model. We extracted dichotomous data indicating the frequency (i.e., number of events and sample size) of values incongruent decisions. To meta-analyze results for decisional conflict, we pooled results using standardized mean differences applying a random effects model. We extracted data on total scores on the Decisional Conflict Scale. We explored and reported consistency using Higgins $I^2$ [42]. We used the Cochrane Risk of Bias tool to assess study bias along 7 domains as well as to assess an overall risk
of bias. Where data permitted, we conducted subgroup meta-analyses of different types of explicit values clarification methods and of explicit values clarification methods that do and do not contain specific design features already identified in previous work [11], namely, whether the method explicitly requires the user to engage with tradeoffs, whether it explicitly provides the user with the implications of what they value, and which, if any, theoretical or conceptual framework underpins it. We used $p=0.05$ as a threshold for statistical significance and conducted analyses in RevMan, version 5.4.

Results

Articles identified

Out of 30,648 records screened at the title and abstract stage and 279 screened at the full text stage, we identified 33 articles that met our inclusion criteria describing trials of 43 values clarification methods. Twenty-four of the articles were new articles identified in this update of IPDAS. We excluded 2 of the articles previously included in the IPDAS Values Clarification Chapter because they did not meet our revised inclusion criteria requiring randomized controlled trials and instead reported, for example, pre-post study designs. The PRISMA diagram of included articles is shown in Figure 1.

Figure 1. PRISMA Diagram
The decision context varied across studies. Out of the 43 included trials, 25 (58%) addressed treatment decisions, 9 (21%) screening decisions, 4 (9%) prevention, 3 (7%) genetic testing, and 2 (5%) diagnostic testing. Thirteen of the 43 trials (30%) centered around a yes/no decision to take an option or not, 18 (42%) a choice between two or more options and 12 (28%) both a yes/no and a choice between two or more options. Most decisions (22/43, 51%) were real decisions, meaning that the person was making this decision in their actual life. The rest were hypothetical (18/43, 42%) or it was not entirely clear whether the decision was real or hypothetical (3/43, 3%). The most commonly-reported outcomes were decisional conflict and/or its subscales (29/43, 67%), decision and/or decisional intentions (22/43, 51%), knowledge (13/43, 30%), and values congruence (12/43, 28%).

As shown in the overview of included studies in Table 3, there was substantial diversity in the types of values clarification methods used. Decision analysis or multicriteria decision analysis was the most commonly trialed method. Full study details are available in Online Appendix 2.

Table 3. Study Details

| Type(s) of Values Clarification Method(s) | Study | Population* | Decision | Summary of findings |
|----------------------------------------|-------|-------------|----------|---------------------|
| Adaptive Conjoint Analysis             | de Achaval and others 2012 [43] | n=208 people with knee osteoarthritis | Whether to receive medication and therapy or total knee arthroplasty | Values clarification method decreased decisional conflict and required more intense cognitive involvement. |
| Adaptive Conjoint Analysis             | Fraenkel and others 2007 [44] | n=87, age at least 60 years old, self report of pain involving one or both knees on most days of the month | Choice between five treatments for knee pain | Values clarification method increased self-confidence in and preparation for shared decision-making, and increased arthritis self-efficacy. |
| Adaptive Conjoint Analysis             | Hess and others 2015 [45] | n=374 women aged 18 years or older with abnormal uterine bleeding & potential candidates for either surgical or medical treatment | Whether or not to be treated for abnormal uterine bleeding, and, if yes, which treatment to undertake | Values clarification method did not reduce decision regret nor improve treatment satisfaction. |
| Adaptive Conjoint Analysis             | Hutyra and others 2019 [46] | n=200 people between 18 and 35 years of age at risk for experiencing a first-time anterior shoulder dislocation | Operative or nonoperative treatment for first-time anterior shoulder dislocation | Values clarification method increased alignment between patients’ treatment decisions and evidence-based recommendations. |
| Adaptive Conjoint Analysis             | Jayadevappa and others 2015 [47] | n=743 people with newly diagnosed localized prostate cancer | Choice between six options for early stage prostate cancer | Values clarification method improved satisfaction with care, satisfaction with decision, reduced regrets, and aligned treatment choice with risk category. |
| Allocation of Points                   | Witteman and others 2020 [48] | n=817 adults asked to imagine they had been diagnosed with colon cancer | Choice between two hypothetical surgeries for colon cancer | Values clarification method (strategy 6b in paper) increased values congruence and reduced decisional conflict. |
| Analytical Hierarchy Process           | Myers and others 2003 [49] | n=199 men aged 50 to 69 with no personal history of prostate cancer/benign prostate hyperplasia | Whether or not to be screened for prostate cancer | Values clarification method decreased rates of prostate cancer screening. Race/ethnicity analyses showed African American men increased screening while white men decreased screening. |
| Analytical Hierarchy Process | Myers and others 2005 [50] | n=242 African-American men, 40-69 years of age and no history of prostate cancer | Whether or not to be screened for prostate cancer, and if yes, choice of method/extent of screening | Values clarification method increased prostate cancer screening. |
|-----------------------------|-----------------------------|-----------------------------------------------------------------|---------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Best-Worst Scaling          | Shirk and others 2017 [51]  | n=122 men with incident localized prostate cancer                | Choice between three options for incident localized prostate cancer              | Values clarification method decreased decisional conflict.                                      |
| Decision Analysis**         | Bekker and others 2004 [52] | n=106 pregnant women receiving a screen positive maternal serum screening result | Whether or not to have a prenatal diagnosis for Down syndrome                    | Values clarification method helped women make more informed prenatal diagnosis decisions.         |
| Decision Analysis**         | Clancy and others 1988 [53] | n=1280 resident and faculty physicians unvaccinated against Hepatitis B | Choice between three options to manage risk of hepatitis B                      | Values clarification method resulted in greater action-taking (screening or vaccination.)        |
| Decision Analysis**         | Feldman-Stewart and others 2012 [54] | n=156 people with newly diagnosed prostate cancer                | Choice between more than five main options for early stage prostate cancer      | Values clarification method increased preparation for decision making and decreased decision regret. Decision conflict decreased with and without values clarification method. |
| Decision Analysis**         | Hopkin and others 2019 [55] | n=349 adults asked to imagine that they had to choose a statin   | Choice between five commonly-used statins                                      | Values clarification method reduced decisional conflict and increased levels of preparation for decision making. |
| Decision Analysis**         | Montgomery and others 2003 [56] | n=217 adults aged 30-80 years with newly diagnosed hypertension | Whether or not to start drug therapy for hypertension                           | Values clarification method increased knowledge and reduced total decisional conflict by significantly reducing scores on uninformed, unclear values and unsupported subscales and somewhat reducing scores on uncertainty subscale. Values clarification method did not influence scores on decision quality subscale, nor did it change state anxiety, decision intention, nor ultimate decision. |
| Decision Analysis**         | Montgomery and others 2007 [57] | n=742 pregnant women with one previous lower segment caesarean section | Choice of planned mode of delivery                                             | Values clarification method reduced decisional conflict and increased frequency of having a vaginal birth. |
| Decision Analysis**         | Witteman and others 2015 [58] | n=407 parents who make medical decisions for at least one child aged 6 months to 18 years and whose child had not yet received the flu vaccine | Whether their child would receive a vaccine against influenza this flu season   | Values clarification method had no effect on values congruence. Values clarification method combined with best practices in risk communication increased intentions to vaccinate, particularly among participants who had not had their children vaccinated against influenza in the past 5 years. |
| Decision Analysis**         | Witteman and others 2020*** [48] | n=1731 adults asked to imagine they had been diagnosed with colon cancer deciding between two treatment options | Choice between two hypothetical surgeries for colon cancer                      | Values clarification method (strategies 2a, 2a+2b, 6c, 6b+6c in paper) increased values congruence and reduced decisional conflict this was when measured (strategies 6c, 6b + 6c in paper). |
| Discrete Choice Experiment  | Brenner and others 2014 [59] | n=615 people between the ages of 50 and 75 at average            | Whether or not to be screened for colorectal cancer, and, if yes,              | Values clarification method influenced choice of most important screening test attribute but did not affect unlabeled test |
| Study Type   | Experiment Description                                                                 | Sample Size | Screening Choice | Decision | Results                                                                 |
|-------------|----------------------------------------------------------------------------------------|-------------|------------------|----------|------------------------------------------------------------------------|
| Discrete    | Experiment (Pignone and others 2012 [60]) n=104 adults 48-75 at average risk for colon  |             |                  |          | Values clarification method influenced choice of most important attribute, but did not affect values clarity, intent to be screened. |
| Discrete    | Experiment (Pignone and others 2013 [61]) n=604 men aged 50-70 at average risk of prostate cancer |             |                  |          | Values clarification method slightly reduced choice of dying as the most important attribute and increased unlabeled PSA-like screening option but did not influence intent to be screened. |
| Open       | Discussion (Au and others 2012 [62]) n=306 people with chronic obstructive pulmonary disease |             | Preferences for end-of-life care |          | Values clarification method helped identify what mattered to patients regarding end-of-life care and communication. Quality of communication improved. |
| Open       | Discussion (Epstein and others 2018 [63]) n=99 people with advanced gastrointestinal cancer |             | Choice between options for end-of-life care |          | Values clarification method improved communication about future medical cancer care but had no effect on decisional conflict or well-being and increased distress. |
| Open       | Discussion (Kennedy and others 2002 [64]) n=894 women with uncomplicated menorrhagia |             | Choice between treatment options for menorrhagia |          | Values clarification method resulted in minimal improvements in self-reported health status, lower use of a more invasive treatment, higher patient satisfaction, more frequent clinician perceptions of "longer than usual" consultations, and lower overall costs. Providing information alone did not affect treatment choices. |
| Open       | Discussion (Lerman and others 1997 [65]) n=700, Women aged 18-75 years who had had at least one first-degree relative with breast and/or ovarian cancer |             | Whether or not to provide a blood sample for BRCA1 testing in the future |          | Values clarification method increased the perceived importance of the limitations and risk of BRCA1 testing and decreased the perceived importance of the benefits of BRCA1 testing. No effect of values clarification method on intent to treat. |
| Open       | Discussion (Matheis-Kraft and others 1997 [66]) n=60 women over age 70 with at least one family member or friend who might act as their proxy to make decisions about life-sustaining treatment |             | Preferences for care in case of decisional incapacity |          | Values clarification method’s effectiveness or lack thereof depended on which statistic (kappa or percent agreement) was used to measure concordance between women and proxies. |
| Pros and Cons | Abhyankar and others 2010 [67] n=30 healthy women asked to imagine having been diagnosed with breast cancer, undergone lumpectomy and suggested chemotherapy by their doctor |             | Choice between having standard adjuvant chemotherapy or taking part in a clinical trial testing a new chemotherapy for early stage breast cancer |          | Values clarification method resulted in more use of personal values when evaluating attributes of options, somewhat less ambivalence, less uncertainty and did not change preferred option. |
| Pros and Cons | O’Connor and others 1999 [68] n=201 women aged 50-69 who had never used hormone therapy |             | Whether or not to take hormone replacement therapy after menopause |          | Values clarification method had no effect on clarity of values, values congruence, total decisional conflict, other subscales of Decisional Conflict Scale, nor acceptability of intervention. |
| Pros and Cons | Paquin and others 2018 [69] n=1000 people aged 18-44 who were pregnant or whose |             | Whether or not to use genomic sequencing to identify genetic |          | Values clarification method decreased parental beliefs against genomic sequencing. |
| Measures of Decisional Conflict | Pros and Cons | Participants | Description | Results |
|--------------------------------|---------------|--------------|-------------|---------|
| Decisional conflict, perceived values clarity | Peinado and others 2020 [10] | n=1000 people aged 18-44 who were pregnant or whose partner was pregnant or planning to become pregnant in the next 2 years | Whether or not to enroll their newborn child in a medical research study that would involve screening for genetic conditions | Values clarification method decreased decisional regret and increased clarity of personal values but had no effect on overall decisional conflict nor on intent to have one's child tested. |
| Decisional conflict, perceived values clarity | Witteman and others 2020 [48] | n=772 adults asked to imagine they had been diagnosed with colon cancer | Choice between two hypothetical surgeries for colon cancer | Values clarification method (strategy 4b in paper) reduced decisional conflict but did not change values congruence. |
| Decisional conflict, perceived values clarity | Garvelink and others 2014*** [70] | n=271 healthy women | Whether or not to undergo fertility preserving procedures prior to cancer treatment | Values clarification method had no effect on knowledge nor decisional conflict. |
| Decisional conflict, perceived values clarity | Kuppermann and others 2014 [71] | n=710 pregnant women who had not yet undergone screening or diagnostic testing for fetal aneuploidy in the current pregnancy | Whether or not to have any screening or diagnostic testing for fetal aneuploidy; if screening or testing is desired, whether to start with screening or with invasive diagnostic testing; and which specific screening and/or diagnostic test(s) to undergo. | Values clarification method increased patient knowledge and resulted in less invasive prenatal test use and more informed choices. Values clarification method did not change decisional conflict nor decisional regret. |
| Decisional conflict, perceived values clarity | Feldman-Stewart and others 2006 [72] | n=90 male volunteers asked to imagine that they had just been diagnosed with early-stage prostate cancer | Choice between four options for early stage prostate cancer | Participants preferred values clarification method with decision analytic summary over values clarification method without summary and no values clarification method. |
| Decisional conflict, perceived values clarity | Witteman and others 2020 [48] | n=785 adults asked to imagine they had been diagnosed with colon cancer | Choice between two hypothetical surgeries for colon cancer | Values clarification method (strategy 6a in paper) reduced decisional conflict but did not change values congruence. |
| Decisional conflict, perceived values clarity | Brenner and others 2014 [59] | n=614 people between the ages of 50 and 75 at average risk for colorectal cancer | Whether or not to be screened for colorectal cancer, and, if yes, what screening test to use | Values clarification method increased the importance placed on risk reduction as an important attribute but did not affect unlabeled test preference, values clarity, nor intent to be screened. |
| Decisional conflict, perceived values clarity | Pignone and others 2012 [60] | n=104 adults 48-75 at average risk for colon cancer | Whether or not to be screened for colorectal cancer, and, if yes, what screening test to use | Values clarification method influenced choice of most important attribute, but did not affect values clarity, intent to be screened, or choice of screening test. |
| Decisional conflict, perceived values clarity | Pignone and others 2013 [61] | n=609 men aged 50-70 at average risk of prostate cancer | Whether or not to be screened for prostate cancer | Values clarification method increased the importance placed on dying but did not influence intent to be screened. |
| Decisional conflict, perceived values clarity | Sheridan and others 2010 [73] | n=137 men aged 45-80 with no prior history of cardiovascular | Whether or not to initiate behaviours to prevent coronary heart disease, and, if so, | Values clarification method had no effect. Decisional conflict, perceived values congruence, and self-efficacy for health behaviours improved with and without |
Quality assessment

Overall study quality was acceptable, with the majority of studies at low risk of bias on most elements. Eight studies were deemed to be at high risk of bias on one element; the majority in Blinding of Participants and Personnel (Performance Bias). Eighteen additional studies were deemed unclear on this element. Blinding of Outcome Assessment (Detection Bias) was the next most common source of potential bias, with 1 study at high risk of bias and 20 more unclear. Full details of risk of bias assessments are available in Online Appendix 3.

Values Congruence

As shown in Figure 2a, included explicit values clarification methods, as a group, increased values congruence. Eleven out of 43 trials (26%) reported the number of people who made values-congruent or values-disgruent decisions. The pooled risk difference of making a values-disgruent decision when using one of the trialed values clarification methods was -0.04 95% CI [-0.06 to -0.02], p<.001. The I² of 28% indicates a relatively low level of statistical heterogeneity.

Figure 2b shows a statistically significant subgroup difference by type of values clarification method. The results suggest that decision analysis is more likely to encourage values-congruent decisions compared to other explicit values clarification methods within this set of trials (Chi-squared(2)=9.25, p=.01). The results show no significant subgroup differences by trade-offs, implementation, theory or by implication of the decision. (See Online Appendix 3.) There were no studies in this analysis with a high risk of bias.
Figure 2a. Values Congruence*: Overall (All Values Clarification Methods Together)

| Study or Subgroup | VCM Events | Total | no VCM Events | Total | Weight | Risk Difference M-H Random, 95% CI Year | Risk Difference M-H Random, 95% CI |
|-------------------|------------|-------|---------------|-------|--------|---------------------------------|---------------------------------|
| O'Connor 1999     | 35         | 101   | 33 100        | 2.6%  | 0.02 [-0.11, 0.15] 1999          |                                  |
| Shandian 2010     | 3          | 62    | 6 75          | 5.9%  | -0.03 [-0.11, 0.05] 2010         |                                  |
| Jayadevappa 2015  | 17         | 343   | 21 350        | 16.7% | -0.01 [-0.04, 0.02] 2015         |                                  |
| Wittenman 2015    | 96         | 205   | 110 202       | 4.4%  | -0.08 [-0.18, 0.02] 2015         |                                  |
| Wittenman 2020 4b | 74         | 350   | 93 422        | 6.9%  | -0.01 [-0.07, 0.05] 2020         |                                  |
| Wittenman 2020 6a | 65         | 393   | 93 422        | 8.8%  | 0.01 [-0.05, 0.07] 2020          |                                  |
| Wittenman 2020 6b | 65         | 395   | 93 422        | 10.9% | -0.05 [-0.19, 0.09] 2020         |                                  |
| Wittenman 2020 6b + 6c | 59 | 413   | 93 422          | 11.6% | -0.09 [-0.13, 0.05] 2020        |                                  |
| Wittenman 2020 6c | 79         | 440   | 93 422        | 11.8% | -0.04 [-0.09, 0.01] 2020         |                                  |
| Wittenman 2020 2a | 43         | 239   | 59 217        | 6.5%  | -0.09 [-0.17, 0.00] 2020         |                                  |
| Wittenman 2020 2a + 2b | 26 | 237   | 41 217        | 6.4%  | -0.08 [-0.14, 0.00] 2020         |                                  |

Total (95% CI) 3148 3271 100.0% -0.04 [-0.06, -0.02]

Total events 564 735
Heterogeneity: $\text{Ta}^2 = 0.00; \text{Ch}^2 = 13.88; \text{df} = 10 (P = 0.18); \chi^2 = 28$
Test for overall effect: $Z = 3.42 (P = 0.0006)$

*Events refer to values-disgruent decisions. The meta-analysis synthesizes the risk across trials of making a values-disgruent decision.

Figure 2b. Values Congruence* by Type of Values Clarification Method

| Study or Subgroup | VCM Events | Total | no VCM Events | Total | Weight | Risk Difference M-H Random, 95% CI Year | Risk Difference M-H Random, 95% CI |
|-------------------|------------|-------|---------------|-------|--------|---------------------------------|---------------------------------|
| 2.2.1 Decision analysis |
| Wittenman 2015    | 96         | 205   | 110 202       | 4.4%  | -0.08 [-0.13, -0.03] 2015         |                                  |
| Wittenman 2020 6b + 6c | 59 | 413   | 93 422          | 11.6% | -0.08 [-0.13, -0.03] 2020        |                                  |
| Wittenman 2020 6c | 76         | 440   | 93 422        | 11.2% | -0.04 [-0.09, 0.01] 2020         |                                  |
| Wittenman 2020 2a | 43         | 239   | 59 217        | 6.5%  | -0.09 [-0.17, -0.02] 2020        |                                  |
| Wittenman 2020 2a + 2b | 26 | 237   | 41 217        | 6.4%  | -0.08 [-0.14, -0.01] 2020        |                                  |
| Subtotal (95% CI) | 1531       | 1480  | 1421          | 42.1% | -0.07 [-0.10, -0.04]             |                                  |

Total events 303 386
Heterogeneity: $\text{Ta}^2 = 0.00; \text{Ch}^2 = 1.67; \text{df} = 4 (P = 0.80); \chi^2 = 0$
Test for overall effect: $Z = 4.79 (P = 0.00001)$

| 2.2.2 Pros and cons |
|-------------------|------------|-------|---------------|-------|--------|---------------------------------|---------------------------------|
| O'Connor 1999     | 35         | 101   | 33 100        | 2.6%  | 0.02 [-0.11, 0.15] 1999          |                                  |
| Wittenman 2020 4b | 74         | 350   | 93 422        | 9.9%  | -0.01 [-0.07, 0.05] 2020         |                                  |
| Subtotal (95% CI) | 451        | 522   | 1221          | 12.5% | -0.00 [-0.06, 0.05]              |                                  |

Total events 100 126
Heterogeneity: $\text{Ta}^2 = 0.00; \text{Ch}^2 = 0.12; \text{df} = 1 (P = 0.73); \chi^2 = 0$
Test for overall effect: $Z = 0.17 (P = 0.88)$

| 2.2.3 Other |
|-------------------|------------|-------|---------------|-------|--------|---------------------------------|---------------------------------|
| Shandian 2010     | 3          | 62    | 6 75          | 6.9%  | -0.03 [-0.11, 0.05] 2010         |                                  |
| Jayadevappa 2015  | 17         | 343   | 21 350        | 16.7% | -0.01 [-0.04, 0.02] 2015         |                                  |
| Wittenman 2020 6a | 65         | 383   | 92 422        | 9.8%  | 0.01 [-0.05, 0.07] 2020          |                                  |
| Wittenman 2020 6b | 66         | 395   | 93 422        | 10.9% | -0.05 [-0.19, 0.09] 2020         |                                  |
| Subtotal (95% CI) | 1193       | 1209  | 1245          | 45.4% | -0.02 [-0.04, 0.00]              |                                  |

Total events 172 213
Heterogeneity: $\text{Ta}^2 = 0.00; \text{Ch}^2 = 2.58; \text{df} = 3 (P = 0.46); \chi^2 = 0$
Test for overall effect: $Z = 1.27 (P = 0.21)$

Total (95% CI) 3148 3271 100.0% -0.04 [-0.06, -0.02]

Total events 564 735
Heterogeneity: $\text{Ta}^2 = 0.00; \text{Ch}^2 = 13.88; \text{df} = 10 (P = 0.18); \chi^2 = 28$
Test for overall effect: $Z = 3.42 (P = 0.0006)$
Test for subgroup differences: $\text{Ta}^2 = 9.25; \text{df} = 2 (P = 0.010); \chi^2 = 78.4$

*Events refer to values-disgruent decisions. The meta-analysis synthesizes the risk across trials of making a values-disgruent decision.
Decisional Conflict

As shown in Figure 3a, explicit values clarification methods decrease decisional conflict. For the 14/43 (33%) trials for which we had complete data, the pooled standardized mean difference for decisional conflict was -0.20 95% CI [-0.29 to -0.11], p<0.001. The I² of 67% represents moderate to high statistical heterogeneity.

Figure 3b shows there was no significant subgroup difference by type of values clarification method (Chi-squared(2)=6.08, p=.05). We found no significant subgroup differences by trade-offs, implementation, theory, implication, nor risk of bias (see Online Appendix 3).

Figure 3a. Decisional Conflict: Overall Measure

| Study or Subgroup | VCM | Mean | SD  | Total | Mean | SD | Total | Weight | IV, Random, 95% CI | Year |
|-------------------|-----|------|-----|-------|------|----|-------|--------|-------------------|------|
| Montgomery 2003   | 27.0| 128  | 100 | 30.8  | 18.3 | 112| 5.7%  |        | -.072 [-1.00, -.44] | 2003 |
| Montgomery 2007   | 23.6| 15.1 | 198 | 27.8  | 14.6 | 201| 7.7%  |        | -.086 [-0.48, -.09] | 2007 |
| Frosh 2008        | 27.95| 13.45| 253 | 30  | 10.77 | 151| 6.9%  |        | -.17 [-0.39, 0.05] | 2008 |
| Garavelik 2014 a  | 41.5| 9.5  | 86  | 44.1 | 12.3 | 55 | 4.4%  |        | -.23 [-0.57, 0.12] | 2014 |
| Kuppermann 2014   | 12.9| 14.1 | 357 | 13.8 | 15.6 | 353| 9.1%  |        | -.06 [-0.21, 0.09] | 2014 |
| Garavelik 2014 b  | 43.0| 14.2 | 70  | 40.9 | 11.6 | 70 | 4.6%  |        | 0.21 [-0.13, 0.54] | 2014 |
| Shrink 2017       | 2.5 | 0.9  | 59  | 2.2 | 0.9 | 63 | 4.2%  |        | 0.32 [-0.33, 0.63] | 2017 |
| Epstein 2018      | 9.9 | 2.35 | 61  | 9.48 | 2.23 | 61 | 4.9%  |        | -.17 [-0.52, 0.19] | 2018 |
| Hopkin 2018       | 14.69| 15.04| 122 | 23.13 | 20.34 | 126| 6.4%  |        | -0.47 [-0.72, -0.22] | 2019 |
| Whiteman 2020 a   | 39  | 27 | 365 | 42 | 27 | 422 | 9.4% |        | -0.11 [-0.25, 0.03] | 2020 |
| Whiteman 2020 b   | 35  | 28 | 350 | 42 | 27 | 422 | 9.2% |        | -0.25 [-0.40, -0.11] | 2020 |
| Whiteman 2020 c   | 36  | 27 | 413 | 42 | 27 | 422 | 9.4% |        | -0.22 [-0.36, -0.09] | 2020 |
| Whiteman 2020 d   | 30  | 26 | 440 | 42 | 27 | 422 | 9.5% |        | -0.23 [-0.36, -0.09] | 2020 |
| Whiteman 2020 e   | 35  | 27 | 363 | 42 | 27 | 422 | 9.3% |        | -0.26 [-0.40, -0.12] | 2020 |

Total (95% CI) 3147 3322 100.0% -0.20 [-0.29, -0.11]

Heterogeneity: Tau² = 0.02; Chi² = 39.50, df = 13 (P = 0.0002), I² = 67%
Test for overall effect: Z = 4.35 (P = 0.0001)
The five studies that compared values clarification methods to each other reported findings that align with the findings of our meta-analyses. Methods that provided users with explicit feedback regarding how the decision options align with their stated values led to somewhat better outcomes, including greater values congruence. [48] When asked to compare methods to each other, study participants also preferred a values clarification method that explicitly showed them how the decision options align with their stated values. [72] Different values clarification methods yielded different patterns of attribute importance. [59–61] Brief summaries of each study are available in Online Appendix 3.

Discussion

Overall, our systematic review and meta-analyses confirm that explicit values clarification methods improve decision outcomes, notably by increasing values congruence and decreasing decisional conflict. Patient decision aids should include an explicit values clarification method.

While the best explicit values clarification method may depend on context—for example, urgent versus routine care or the extent to which a decision has a clear set of decision attributes—our analyses suggest that patient decision aid developers may wish to consider methods that draw on multicriteria decision analysis. The apparent advantages of such methods shown in our analyses may reflect similarities between the process and the outcome. In other words, increased values congruence yielded by decision analytic methods may be a function of the ways in which such methods transparently show people how their options align with their stated values. We additionally caution that when these methods use pre-specified attributes, there might not be the
flexibility for users to add new attributes, highlighting the importance of research to inform attribute selection. We acknowledge that some researchers have argued that health professionals having an unhurried, high-quality conversation with patients may be a preferred approach for at least some patients, especially when decision attributes are many and varied. However, in this systematic review, trials of Open Discussion values clarification methods did not demonstrate strong results, suggesting that such an ideal may be difficult to achieve.

To advance further knowledge on the merits and pitfalls of different values clarification methods, we recommend that authors of future trials of values clarification methods report four outcomes: decisional conflict, decision or decision intention, values congruence, and decisional regret. When possible, authors should make use of validated scales that have good psychometric properties and are commonly reported, as this facilitates evidence synthesis.

Decisonal conflict should be assessed before people make the decision, using a version of the Decisional Conflict Scale. [75, 76] Decisions or decision intentions should be assessed when the decision is made.

Values congruence should be assessed once the decision is made. We acknowledge that including values congruence as an outcome brings both measurement and conceptual issues. Measurement issues exist because there are disagreements about how to measure what matters to people (or indeed, whether it is conceptually possible to do so) and compare such measures to what people choose. [77] Values congruence should not be measured using the values clarity subscale of the Decisional Conflict Scale, as this subscale measures perceived values clarity, not values congruence. [48] Further research is required to determine whether measuring values congruence might introduce bias or otherwise negatively influence decision making.

Decisional regret should be assessed with a version of the Decisional Regret Scale [78, 79] after people make the decision, ideally with a sufficiently long delay that longer-term effects can be captured. An included study in this review showed that a values clarification method reduced decisional regret, but only after a year had passed following implementation of the decision. [54]

For all four measures, authors should clearly report sample mean and sample standard deviation for continuous measures, numbers in each category for categorical measures, and sample size per study arm in all cases. Finally, we recommend that patient decision aid developers explain the rationale for their choice of values clarification method.

Our study has three main limitations. First, the included data were of moderate quality. Although this review includes many robust trials, the included studies often measured different outcomes or the same outcomes in different ways, there were missing data in some studies, some studies had high risk of bias (often because it was not possible to prevent study participants from ascertaining the study arm to which they were assigned), and some of our meta-analyses had high heterogeneity. Together, these issues suggest a degree of caution in our conclusions. Second, we did not distinguish between subtypes of values clarification methods. For example, different adaptive conjoint analysis exercises may be very different from each other, as might open discussions, or many other values clarification methods we grouped together, particularly those we grouped under the broad umbrella term of multicriteria decision analysis. Indeed, the values clarification methods used and trialed may simply reflect authors’ interests and expertise. Given the breadth of methods available, further comparative effectiveness research is needed to conclusively determine the superiority of any given method. Third and finally, our primary findings were heavily influenced by studies conducted with relatively homogenous populations making hypothetical decisions. Although our sensitivity analyses suggested no differences
between studies in real and hypothetical contexts, we nonetheless believe further study is needed in more diverse populations making real decisions before drawing firmer conclusions.

Our study also has three main strengths. First, we catalog definitions and resources regarding values clarification methods, as well as recommended outcomes to report in studies. In doing so, we hope to offer more clarity and structure to a literature that can be confusing to navigate, particularly for those who are newer to developing patient decision aids. Second, we begin to answer a core question that commonly arises when developing a patient decision aid: when including a values clarification method, which type of method should one use? Third and finally, we used rigorous methods and an expansive, systematic search. By conducting a systematic review, we reduced our likelihood of missing relevant studies. By including meta-analyses, we offer stronger findings and recommendations than would be possible without pooling data across multiple studies.

In conclusion, patient decision aids should include an explicit values clarification method. Patient decision aid developers may wish to consider the potential advantages of multicriteria decision analysis. Future research should further investigate which methods lead to the best outcomes across or within particular decisions, populations, and settings. Authors of randomized controlled trials in this area should report decisional conflict, decision made, values congruence, and decisional regret.
DECLARATIONS

Abbreviations
None

Ethics Approval, Consent to Participate and Consent for Publication
Not applicable.

Availability of Data and Materials
Full data are available in the online appendices.

Competing Interests
None.

Funding
This study was funded by the Canadian Institutes of Health Research (CIHR) FDN-148426 (PI Witteman). HOW receives salary support from Tier 2 Canada Research Chair in Human-Centred Digital Health and received salary support during this study from a Fonds de Recherche du Québec-Santé (FRQS) Research Scholar Junior 2 Career Award. The CIHR, Canada Research Chairs program, and FRQS had no role in determining the study design, the plans for data collection or analysis, the decision to publish, nor the preparation of this manuscript.

Authors’ Contributions
All authors contributed to the design of the study. HOW, RN, GV, SCD, SC, MM, TP, CR, CRB, JJ contributed to data collection. HOW, SCD, and JJ conducted data analysis and interpretation. HOW drafted the first version of the article with early revision by RN, GV, SCD, SC, MM, TP, CR, CRB, JJ and multiple subsequent revisions by all authors. All authors critically revised the article and approved the final version for submission for publication. HOW had full access to all the data in the study and had final responsibility for the decision to submit for publication. Authors BA, JFPB, SC, AF, TG, MM, AP, MP, TP, CR, DR, CRB, PT, MW, DBW, CEW contributed approximately equally and are listed alphabetically by last name.

Acknowledgements
The authors gratefully acknowledge Frédéric Bergeron, MLIS, for assistance with search strategy and Caroline Beaudoin for assistance in resolving article counts. We thank all authors of the original articles who generously gave their time to provide missing data when we were unable to extract the data needed from their papers.
References

[1] Marteau TM, Dormandy E, Michie S. A measure of informed choice. *Health Expect* 2001; 4: 99–108.

[2] Llewellyn-Thomas HA. Patients’ health-care decision making: a framework for descriptive and experimental investigations. *Med Decis Making* 1995; 15: 101–106.

[3] Légaré F, Witteman HO. Shared decision making: examining key elements and barriers to adoption into routine clinical practice. *Health Aff* 2013; 32: 276–284.

[4] Hoffmann TC, Montori VM, Del Mar C. The connection between evidence-based medicine and shared decision making. *JAMA* 2014; 312: 1295–1296.

[5] Barrett B, McKenna P. Communicating benefits and risks of screening for prostate, colon, and breast cancer. *Fam Med* 2011; 43: 248–253.

[6] Mulley AG, Trimble C, Elwyn G. Stop the silent misdiagnosis: patients’ preferences matter. *BMJ* 2012; 345: e6572.

[7] Llewellyn-Thomas HA, Crump RT. Decision support for patients: values clarification and preference elicitation. *Med Care Res Rev* 2013; 70: 50S–79S.

[8] Munro S, Stacey D, Lewis KB, et al. Choosing treatment and screening options congruent with values: Do decision aids help? Sub-analysis of a systematic review. *Patient Educ Couns* 2016; 99: 491–500.

[9] Stacey D, Légaré F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. Epub ahead of print 2017. DOI: 10.1002/14651858.CD001431.pub5.

[10] Peinado S, Paquin RS, Rini C, et al. Values clarification and parental decision making about newborn genomic sequencing. *Health Psychol*. Epub ahead of print 30 December 2019. DOI: 10.1037/hea0000829.

[11] Witteman HO, Scherer LD, Gavaruzzi T, et al. Design Features of Explicit Values Clarification Methods: A Systematic Review. *Med Decis Making* 2016; 36: 453–471.

[12] Witteman HO, Gavaruzzi T, Scherer LD, et al. Effects of Design Features of Explicit Values Clarification Methods: A Systematic Review. *Med Decis Making* 2016; 36: 760–776.

[13] Bridges JFP, Brett Hauber A, Marshall D, et al. Conjoint Analysis Applications in Health—a Checklist: A Report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value in Health* 2011; 14: 403–413.

[14] Ho M, Saha A, McCleary KK, et al. A Framework for Incorporating Patient Preferences Regarding Benefits and Risks into Regulatory Assessment of Medical Technologies. *Value Health* 2016; 19: 746–750.

[15] Kristensen FB, Husereau D, Huić M, et al. Identifying the Need for Good Practices in Health Technology Assessment: Summary of the ISPOR HTA Council Working Group Report on Good Practices in HTA. *Value in Health* 2019; 22: 13–20.

[16] O’Connor A, Llewellyn-Thomas H, Dolan J, et al. Section D: Clarifying and Expressing Values. In: O’Connor A, Llewellyn-Thomas H, Stacey D (eds) *IPDAS Collaboration Background Document*. International Patient Decision Aids Standards (IPDAS) Collaboration, 2005, pp. 17–23.

[17] Fagerlin A, Pignone M, Abhyankar P, et al. Clarifying values: an updated review. *BMC Med Inform Decis Mak* 2013; 13 Suppl 2: S8.
[18] Charles C, Gafni A. The vexing problem of defining the meaning, role and measurement of values in treatment decision-making. *J Comp Eff Res* 2014; 3: 197–209.

[19] Lim C, Berry ABL, Hirsch T, et al. Understanding What Is Most Important to Individuals with Multiple Chronic Conditions: A Qualitative Study of Patients’ Perspectives. *J Gen Intern Med* 2017; 32: 1278–1284.

[20] Rocque R, Chipenda Dansokho S, Grad R, et al. What Matters to Patients and Families: A Content and Process Framework for Clarifying Preferences, Concerns, and Values. *Med Decis Making* 2020; 272989X20940660.

[21] Patient Preference Information – Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. Food and Drug Administration, https://www.fda.gov/media/92593/download (24 August 2016).

[22] Pieterse AH, de Vries M, Kunneman M, et al. Theory-informed design of values clarification methods: a cognitive psychological perspective on patient health-related decision making. *Soc Sci Med* 2013; 77: 156–163.

[23] de Vries M, Fagerlin A, Witteman HO, et al. Combining deliberation and intuition in patient decision support. *Patient Educ Couns* 2013; 91: 154–160.

[24] Fischhoff B, Barnato AE. Value Awareness: A New Goal for End-of-life Decision Making. *MDM Policy & Practice* 2019; 4: 2381468318817523.

[25] Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *Br Med Bull* 2012; 103: 21–44.

[26] Whitty JA, Lancsar E, Rixon K, et al. A systematic review of stated preference studies reporting public preferences for healthcare priority setting. *Patient* 2014; 7: 365–386.

[27] Pieterse AH, Berkers F, Baas-Thijssen MCM, et al. Adaptive Conjoint Analysis as individual preference assessment tool: feasibility through the internet and reliability of preferences. *Patient Educ Couns* 2010; 78: 224–233.

[28] Bansback N, Li LC, Lynd L, et al. Development and preliminary user testing of the DCIDA (dynamic computer interactive decision application) for ‘nudging’ patients towards high quality decisions. *BMC Med Inform Decis Mak* 2014; 14: 1–13.

[29] Dolan JG, Boohaker E, Allison J, et al. Patients’ preferences and priorities regarding colorectal cancer screening. *Med Decis Making* 2013; 33: 59–70.

[30] Ungar WJ, Hadioonzadeh A, Najafzadeh M, et al. Quantifying preferences for asthma control in parents and adolescents using best-worst scaling. *Respir Med* 2014; 108: 842–851.

[31] Thokala P, Devlin N, Marsh K, et al. Multiple Criteria Decision Analysis for Health Care Decision Making--An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health* 2016; 19: 1–13.

[32] Regier DA, Peacock S. Theoretical Foundations of MCDA. In: Marsh K, Goetghebeur M, Thokala P, et al. (eds) *Multi-Criteria Decision Analysis to Support Healthcare Decisions*. Cham: Springer International Publishing, 2017, pp. 9–28.

[33] Hazlewood GS, Marshall DA, Barber CEH, et al. Using a Discrete-Choice Experiment in a Decision Aid to Nudge Patients Towards Value-Concordant Treatment Choices in Rheumatoid Arthritis: A Proof-of-Concept Study. *Patient Prefer Adherence* 2020; 14: 829–838.
[34] Weir KR, Bonner C, McCaffery K, et al. Pharmacists and patients sharing decisions about medicines: Development and feasibility of a conversation guide. *Res Social Adm Pharm* 2019; 15: 682–690.

[35] Ottawa Personal Decision Guides - Patient Decision Aids - Ottawa Hospital Research Institute, https://decisionaid.ohri.ca/decguide.html (accessed 4 November 2020).

[36] Masya LM, Young JM, Solomon MJ, et al. Preferences for outcomes of treatment for rectal cancer: patient and clinician utilities and their application in an interactive computer-based decision aid. *Dis Colon Rectum* 2009; 52: 1994–2002.

[37] Heckerling PS, Verp MS, Albert N. Patient or physician preferences for decision analysis: the prenatal genetic testing decision. *Med Decis Making* 1999; 19: 66–77.

[38] Jibaja-Weiss ML, Volk RJ. Utilizing computerized entertainment education in the development of decision aids for lower literate and naive computer users. *J Health Commun* 2007; 12: 681–697.

[39] Stiggelbout AM, Kiebert GM, Kievit J, et al. Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores. *Med Decis Making* 1994; 14: 82–90.

[40] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; 6: e1000097.

[41] Higgins JPT, Green S (eds). *Cochrane handbook for systematic reviews for interventions*, https://handbook-5-1.cochrane.org/front_page.htm (2011).

[42] Higgins JPT, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557–560.

[43] de Achaval S, Fraenkel L, Volk RJ, et al. Impact of educational and patient decision aids on decisional conflict associated with total knee arthroplasty. *Arthritis Care Res* 2012; 64: 229–237.

[44] Fraenkel L, Rabidou N, Wittink D, et al. Improving informed decision-making for patients with knee pain. *J Rheumatol* 2007; 34: 1894–1898.

[45] Hess LM, Litwiller A, Byron J, et al. Preference elicitation tool for abnormal uterine bleeding treatment: a randomized controlled trial. *Patient* 2015; 8: 217–227.

[46] Hutyra CA, Smiley S, Taylor DC, et al. Efficacy of a Preference-Based Decision Tool on Treatment Decisions for a First-Time Anterior Shoulder Dislocation: A Randomized Controlled Trial of At-Risk Patients. *Med Decis Making* 2019; 39: 253–263.

[47] Jayadevappa R, Chhatre S, Gallo JJ, et al. Treatment preference and patient centered prostate cancer care: Design and rationale. *Contemp Clin Trials* 2015; 45: 296–301.

[48] Witteman HO, Julien A-S, Ndjaboue R, et al. What Helps People Make Values-Congruent Medical Decisions? Eleven Strategies Tested across 6 Studies. *Med Decis Making* 2020; 40: 266–278.

[49] Myers RE. *Value-based decision-making in prostate cancer early detection*. DAMD17-98-1-8641, Thomas Jefferson University, November 2003.

[50] Myers RE, Daskalakis C, Cocroft J, et al. Preparing African-American men in community primary care practices to decide whether or not to have prostate cancer screening. *J Natl Med Assoc* 2005; 97: 1143–1154.

[51] Shirk JD, Crespi CM, Saucedo JD, et al. Does Patient Preference Measurement in Decision Aids Improve Decisional Conflict? A Randomized Trial in Men with Prostate Cancer. *Patient*. Epub ahead of print 16
[52] Bekker HL, Hewison J, Thornton JG. Applying decision analysis to facilitate informed decision making about prenatal diagnosis for Down syndrome: a randomised controlled trial. *Prenat Diagn* 2004; 24: 265–275.

[53] Clancy CM, Cebul RD, Williams SV. Guiding individual decisions: a randomized, controlled trial of decision analysis. *Am J Med* 1988; 84: 283–288.

[54] Feldman-Stewart D, Tong C, Siemens R, et al. The Impact of Explicit Values Clarification Exercises in a Patient Decision Aid Emerges After the Decision Is Actually Made. *Medical Decision Making* 2012; 32: 616–626.

[55] Hopkin G, Au A, Collier VJ, et al. Combining Multiple Treatment Comparisons with Personalized Patient Preferences: A Randomized Trial of an Interactive Platform for Statin Treatment Selection. *Med Decis Making* 2019; 39: 264–277.

[56] Montgomery AA, Fahey T, Peters TJ. A factorial randomised controlled trial of decision analysis and an information video plus leaflet for newly diagnosed hypertensive patients. *Br J Gen Pract* 2003; 53: 446–453.

[57] Montgomery AA, Emmett CL, Fahey T, et al. Two decision aids for mode of delivery among women with previous caesarean section: Randomized controlled trial. *BMJ: British Medical Journal* 2007; 334: 1305–1309.

[58] Witteman HO, Chipenda Dansokho S, Exe N, et al. Risk Communication, Values Clarification, and Vaccination Decisions. *Risk Anal* 2015; 35: 1801–1819.

[59] Brenner A, Howard K, Lewis C, et al. Comparing 3 values clarification methods for colorectal cancer screening decision-making: a randomized trial in the US and Australia. *J Gen Intern Med* 2014; 29: 507–513.

[60] Pignone MP, Brenner AT, Hawley S, et al. Conjoint analysis versus rating and ranking for values elicitation and clarification in colorectal cancer screening. *J Gen Intern Med* 2012; 27: 45–50.

[61] Pignone MP, Howard K, Brenner AT, et al. Comparing three techniques for eliciting patient values for decision making about prostate-specific antigen screening: a randomized controlled trial. *JAMA Intern Med* 2013; 173: 362–368.

[62] Au DH, Udris EM, Engelberg RA, et al. A randomized trial to improve communication about end-of-life care among patients with COPD. *Chest* 2012; 141: 726–735.

[63] Epstein AS, O'Reilly EM, Shuk E, et al. A Randomized Trial of Acceptability and Effects of Values-Based Advance Care Planning in Outpatient Oncology: Person-Centered Oncologic Care and Choices. *J Pain Symptom Manage* 2018; 56: 169–177.e1.

[64] Kennedy ADM, Sculpher MJ, Coulter A, et al. Effects of Decision Aids for Menorrhagia on Treatment Choices, Health Outcomes, and Costs: a randomized controlled trial. *JAMA* 2002; 288: 2701.

[65] Lerman C, Biesecker B, Benkendorf JL, et al. Controlled trial of pretest education approaches to enhance informed decision-making for BRCA1 gene testing. *J Natl Cancer Inst* 1997; 89: 148–157.

[66] Matheis-Kraft C, Roberto KA. Influence of a values discussion on congruence between elderly women and their families on critical health care decisions. *J Women Aging* 1997; 9: 5–22.

[67] Abhyankar P, Bekker HL, Summers BA, et al. Why values elicitation techniques enable people to make informed decisions about cancer trial participation. *Health Expect* 2011; 14 Suppl 1: 20–32.
O'Connor RN AM, Wells GA, Fr Cp PTMD, et al. The effects of an ‘explicit’ values clarification exercise in a woman’s decision aid regarding postmenopausal hormone therapy. *Health Expectations* 1999; 21–32.

Paquin RS, Peinado S, Lewis MA, et al. A behavior-theoretic evaluation of values clarification on parental beliefs and intentions toward genomic sequencing for newborns. *Soc Sci Med* 2018; 112037.

Garvelink MM, ter Kuile MM, Stiggelbout AM, et al. Values clarification in a decision aid about fertility preservation: does it add to information provision? *BMC Med Inform Decis Mak* 2014; 14: 68.

Kuppermann M, Pena S, Bishop JT, et al. Effect of enhanced information, values clarification, and removal of financial barriers on use of prenatal genetic testing: a randomized clinical trial. *JAMA* 2014; 312: 1210–1217.

Feldman-Stewart D, Brennenstuhl S, Brundage MD, et al. An explicit values clarification task: development and validation. *Patient Educ Couns* 2006; 63: 350–356.

Sheridan SL, Griffith JM, Behrend L, et al. Effect of adding a values clarification exercise to a decision aid on heart disease prevention: a randomized trial. *Med Decis Making* 2010; 30: E28–39.

Frosch D, Bhatnagar V, Tally S, et al. Internet patient decision support. *Arch Intern Med* 2008; 168: 363–369.

O’Connor AM. *User Manual – Decisional Conflict Scale*. University of Ottawa, https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Decisional_Conflict.pdf (2010).

Decisional conflict scale - evaluation measures - patient decision aids - Ottawa hospital research institute, https://decisionaid.ohri.ca/eval_dcs.html (accessed 20 January 2021).

Winn K, Ozanne E, Sepucha K. Measuring patient-centered care: An updated systematic review of how studies define and report concordance between patients’ preferences and medical treatments. *Patient Educ Couns* 2015; 98: 811–821.

Brehaut JC, O’Connor AM, Wood TJ, et al. Validation of a decision regret scale. *Med Decis Making* 2003; 23: 281–292.

Decision regret scale - evaluation measures - patient decision aids - Ottawa hospital research institute, https://decisionaid.ohri.ca/eval_regret.html (accessed 20 January 2021).
Online Appendix 1: Search Strategies

#1 ((value? OR patient preference? OR treatment preference?) adj5 (clarif* OR elicit*).ti,ab,kw
#2 (Decision Making/ AND Social Values/)
#3 MCDA.ti,ab
#4 analytical hierarchy process.ti,ab,kw
#5 best-worst scaling.ti,ab,kw
#6 ((conjoint OR decision) adj3 analysis):ti,ab,kw
#7 data envelopment analysis.ti,ab,kw
#8 Decision conferencing.ti,ab,kw
#9 Decision models.ti,ab,kw
#10 direct rating.ti,ab,kw
#11 points allocation.ti,ab,kw
#12 discrete choice experiment.ti,ab,kw
#13 ("dominance-based" adj3 approach*).ti,ab,kw
#14 EVIDEM framework.ti,ab,kw
#15 (geometrical analysis for interactive aid OR GAIA).ti,ab,kw
#16 MACBETH.ti,ab,kw
#17 ("Measuring Attractiveness" adj4 "Categorical Based Evaluation TecHnique").ti,ab,kw
#18 Multi-Attribute Global Inference of Quality.ti,ab,kw
#19 ("Multiple attribute" OR "multiple criteria" OR multiattribute) adj2 (utility OR theory OR analysis)).ti,ab,kw
#20 (MAUT OR MAVT OR MCUA OR MCA).ti,ab,kw
#21 ("Novel approach to imprecise assessment and decision environments" OR NAIADE).ti,ab,kw
#22 ORESTE.ti,ab,kw
#23 Pairwise comparisons.ti,ab,kw
#24 PAPRIKA.ti,ab,kw
#25 Pairwise RanKings.ti,ab,kw
#26 PROMETHEE.ti,ab,kw
#27 Preference Ranking Organization Method for Enrichment of Evaluations.ti,ab,kw
#28 QUALitative FLEXible.ti,ab,kw
#29 Simple Multi Attribute Rating Technique.ti,ab,kw
#30 SMART.ti,ab,kw
#31 Standard gamble.ti,ab,kw
#32 Swing weighting.ti,ab,kw
#33 TOPSIS.ti,ab,kw
Technique for Order Preference by Similarity to the Ideal Solution. 
(Time tradeoff OR time tradeoff). 
Value function methods. 
Valutazione delle Tecnologie Sanitarie. 
VDA. 
VTS. 
verbal decision analysis. 
visual analog scale. 
willingness-to-pay. 
((Scoring OR weighting) adj1 methods). 
REGIME. 
(scal* adj2 (methods OR Natural OR Constructed OR Objective)). 
OR/3-45 
Randomized Controlled Trials as Topic/ 
randomized controlled trial/ 
Random Allocation/ 
Double Blind Method/ 
Single Blind Method/ 
clinical trial/ 
clinical trial, phase i.pt 
clinical trial, phase ii.pt 
clinical trial, phase iii.pt 
clinical trial, phase iv.pt 
controlled clinical trial.pt 
randomized controlled trial.pt 
multicenter study.pt 
clinical trial.pt 
expc Clinical Trials as topic/ 
(clinical adj trial$).tw 
\((\text{singl}\|$ \text{or doubl}\|$ \text{or treb}\|$ \text{or tripl}\$) \text{adj (blind$3$ or mask$3$)})\).tw 
PLACEBOS/ 
placebo$.tw 
randomly allocated.tw
(allocated adj2 random$).tw
OR/47-67
(#1 OR #2 OR #46) AND #68
#70 NOT (animals/ NOT humans/)
(((value* OR "patient preference" OR "treatment preferences") NEAR/5 (clarif* OR elicit*)):ti,ab,kw)

'decision support system'/de OR 'patient decision making'/de

MCDA:ti,ab,kw

"analytical hierarchy process":ti,ab,kw

"best-worst scaling":ti,ab,kw

((conjoint OR decision) NEAR/3 analysis):ti,ab,kw

"data envelopment analysis":ti,ab,kw

"Decision conferencing":ti,ab,kw

"Decision models":ti,ab,kw

"direct rating":ti,ab,kw

"points allocation":ti,ab,kw

"discrete choice experiment":ti,ab,kw

("dominance-based" NEAR/3 approach*):ti,ab,kw

"EVIDEM framework":ti,ab,kw

"geometrical analysis for interactive aid OR GAIA":ti,ab,kw

MACBETH:ti,ab,kw

("Measuring Attractiveness" NEAR/4 "Categorical Based Evaluation Technique"):ti,ab,kw

"Multi-Attribute Global Inference of Quality":ti,ab,kw

(("Multiple attribute" OR "multiple criteria" OR multiattribute) NEAR/2 (utility OR theory OR analysis)):ti,ab,kw

(MAUT OR MAVT OR MCUA OR MCA):ti,ab,kw

("Novel approach to imprecise assessment and decision environments" OR NAIADE):ti,ab,kw

ORESTE:ti,ab,kw

"Pairwise comparisons":ti,ab,kw

PAPRIKA:ti,ab,kw

"Pairwise RanKings":ti,ab,kw

PROMETHEE:ti,ab,kw

"Preference Ranking Organization Method for Enrichment of Evaluations":ti,ab,kw

"QUALitative FLEXible":ti,ab,kw

"Simple Multi Attribute Rating Technique":ti,ab,kw

SMART:ti,ab,kw

"Standard gamble":ti,ab,kw

"Swing weighting":ti,ab,kw

TOPSIS:ti,ab,kw
"Technique for Order Preference by Similarity to the Ideal Solution":ti,ab,kw
("Time tradeoff" OR "time tradeoff"):ti,ab,kw
"Value function methods":ti,ab,kw
"Valutazione delle Tecnologie Sanitarie":ti,ab,kw
VDA:ti,ab,kw
VTS:ti,ab,kw
"verbal decision analysis":ti,ab,kw
"visual analog scale":ti,ab,kw
"willingness-to-pay":ti,ab,kw
((Scoring OR weighting) NEAR/1 methods):ti,ab,kw
REGIME:ti,ab,kw
(scal* NEAR/2 (methods OR Natural OR Constructed OR Objective)):ti,ab,kw
#3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR i
'clinical trial'/de
'randomized controlled trial'/de
'randomization'/de
'single blind procedure'/de
'double blind procedure'/de
'crossover procedure'/de
'placebo'/de OR 'prospective study'/de
(randomized NEXT/1 controlled NEXT/1 trial):ab,ti
rct:ab,ti OR 'random allocation':ab,ti
'randomly allocated':ab,ti OR 'allocated randomly':ab,ti
(allocated NEAR/2 random):ab,ti
((single OR double OR treble OR triple) NEXT/1 blind*):ab,ti
placebo*:ab,ti
#47 OR #48 OR #49 OR #50 OR #51 OR #52 OR #53 OR #54 OR #55 OR #56 OR #57 OR #58 OR #59
(#1 OR #2 OR #46) AND #60
#61 NOT (animal:de NOT human:de)
[embase]/lim NOT ([embase]/lim AND [medline]/lim)
#1 ((value* OR "patient preference*" OR "treatment preference*") NEAR/5 (clarify* OR elicit*)):ti,ab
#2 ("Decision Making":kw AND "Social Values":kw)
#3 MCDA:ti,ab
#4 "analytical hierarchy process":ti,ab
#5 best-worst scaling:ti,ab
#6 ((conjoint OR decision) NEAR/3 analysis):ti,ab
#7 "Decision conferencing":ti,ab
#8 "Decision models":ti,ab
#9 "direct rating":ti,ab
#10 "points allocation":ti,ab
#11 "discrete choice experiment":ti,ab
#12 ("dominance-based" NEAR/3 approach*):ti,ab
#13 "Elimination and Choice Expressing Reality":ti,ab
#14 "EVIDEM framework":ti,ab
#15 "geometrical analysis for interactive aid" OR GAIA:ti,ab
#16 MACBETH:ti,ab
#17 ("Measuring Attractiveness" NEAR/4 "Categorical Based Evaluation Technique"):ti,ab
#18 "Multi-Attribute Global Inference of Quality":ti,ab
#19 "multiattribute objective function specification":ti,ab
#20 ("Multiple attribute" OR "multiple criteria" OR multiattribute) NEAR/2 (utility OR theory OR analysis)):ti,ab
#21 (MAUT OR MAVT OR MCUA OR MCA):ti,ab
#22 ("Novel approach to imprecise assessment and decision environments" OR NAIADE):ti,ab
#23 ORESTE:ti,ab
#24 "Pairwise comparisons":ti,ab
#25 PAPRIKA:ti,ab
#26 "Pairwise RanKings":ti,ab
#27 PROMETHEE:ti,ab
#28 "Preference Ranking Organization Method for Enrichment of Evaluations":ti,ab
#29 "QUALitative FLEXible":ti,ab
#30 "Simple Multi Attribute Rating Technique":ti,ab
#31 SMART:ti,ab
#32 "Standard gamble":ti,ab
#33 "Swing weighting":ti,ab
TOPSIS:ti,ab
"Technique for Order Preference by Similarity to the Ideal Solution":ti,ab
("Time tradeoff" OR "time tradeoff"):ti,ab
"Value function methods":ti,ab
"Valutazione delle Tecnologie Sanitarie":ti,ab
VDA:ti,ab
VTS:ti,ab
"verbal decision analysis":ti,ab
"visual analog scale":ti,ab
"willingness-to-pay":ti,ab
scoring method*:ti,ab OR "weighting method**":ti,ab
REGIME:ti,ab
(scal* NEAR/2 (methods OR Natural OR Constructed OR Objective)):ti,ab
{OR #3-#46}
#1 OR #2 OR #47
#1 TS=((value OR "patient preference" OR "treatment preference") NEAR/5 (clarif* OR elicit*)))
#2 TS=((conjoint OR decision) NEAR/3 analysis)
#3 TS=(MCDA)
#4 TS=('analytical hierarchy process')
#5 TS=('best-worst scaling')
#6 TS=('data envelopment analysis')
#7 TS=('Decision conferencing')
#8 TS=('Decision models')
#9 TS=('direct rating')
#10 TS=('points allocation')
#11 TS=('discrete choice experiment')
#12 TS=('dominance-based' NEAR/3 approach*)
#13 TS=('Elimination and Choice Expressing Reality')
#14 TS=('EVIDEM framework')
#15 TS=('geometrical analysis for interactive aid' OR GAIA)
#16 TS=(MACBETH)
#17 TS=('Measuring Attractiveness' NEAR/4 'Categorical Based Evaluation TechNique')
#18 TS=('Multi-Attribute Global Inference of Quality')
#19 TS=('Multiple attribute' OR 'multiple criteria' OR multiattribute) NEAR/2 (utility OR theory OR analysis))
#20 TS=(MAUT OR MAVT OR MCUA OR MCA)
#21 TS=('Novel approach to imprecise assessment and decision environments' OR NAIADE)
#22 TS=(ORESTE)
#23 TS=('Pairwise comparisons')
#24 TS=(PAPRIKA)
#25 TS=('Pairwise RanKings')
#26 TS=(PROMETHEE)
#27 TS=('Preference Ranking Organization Method for Enrichment of Evaluations')
#28 TS=('QUALitative FLEXible')
#29 TS=('Simple MultiAttribute Rating Technique')
#30 TS=(SMART)
#31 TS=('Standard gamble')
#32 TS=(TOPSIS)
#33 TS=('Technique for Order Preference by Similarity to the Ideal Solution')
#34 TS="("time tradeoff" OR "time tradeoff")"
#35 TS="("Value function methods")"
#36 TS="("Valutazione delle Tecnologie Sanitarie")"
#37 TS=(VDA)
#38 TS=(VTS)
#39 TS="("verbal decision analysis")"
#40 TS="("visual analog scale")"
#41 TS="("willingness-to-pay")"
#42 TS= ((Scoring OR weighing) NEAR/1 methods)
#43 TS=(REGIME)
#44 TS=(scal* NEAR/2 (methods OR Natural OR Constructed OR Objective))
#45 #44 OR #43 OR #42 OR #41 OR #40 OR #39 OR #38 OR #37 OR #36 OR #35 OR #34 OR #33 OR #32 OR #31 OR #30 OR #29 OR #28 OR #27 OR #26 OR #25 OR #24 OR #23 OR #22 OR #21 OR #20 OR #19 OR #18 OR #17 OR #16 OR #15 OR #14 OR #13 OR #12 OR #11 OR #10 OR #9 OR #8 OR #7 OR #6 OR #5 OR #4 OR #3 OR #2
#46 "clinical trial"
#47 randomization
#48 "crossover procedure"
#49 placebo
#50 "prospective study"
#51 (randomized NEAR/1 "controlled trial")
#52 rct
#53 (allocated NEAR/2 random*)
#54 ((single OR double OR treble OR triple) NEAR/1 blind*)
#55 #46 OR #47 OR #48 OR #49 OR #50 OR #51 OR #52 OR #53 OR #54
#56 (#1 OR #45) AND #55
#34  TI VTS OR AB VTS
#35  TI "visual analog scale" OR AB "visual analog scale"
#36  TI "willingness-to-pay" OR AB "willingness-to-pay"
#37  TI Scoring OR "weighting method#" OR AB Scoring OR "weighting method#"
#38  TI REGIME OR AB REGIME
#39  TI (scal* N2 (methods OR Natural OR Constructed OR Objective)) OR AB (scal* N2 (methods OR Natural OR Constructe
#40  S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR
#41  TX allocat* random*
#42  MH "Quantitative Studies"
#43  MH "Placebos"
#44  TX placebo*
#45  TX random* allocat*
#46  MH "Random Assignment"
#47  TX randomi* control* trial*
#48  TX ((singl* n1 blind*) OR (singl* n1 mask*))
#49  TX ((doubl* n1 blind*) OR (doubl* n1 mask*))
#50  TX ((tripl* n1 blind*) OR (tripl* n1 mask*))
#51  TX ((trebl* n1 blind*) or (trebl* n1 mask*))
#52  TX (clinic* n1 trial*)
#53  PT "Clinical trial"
#54  MH "Clinical Trials+"
#55  #52 OR #65
#56  (#1 OR #2 OR #3 OR #4 OR #40) AND #55
### Summary of findings

Participants were randomized to one of four groups: (1) traditional consultations, (2) random permuted block to structured information plus VCM, (3) information plus VCM, and (4) video alone. The results suggest that decision aids with or without values elicitation techniques lead to improved prostate cancer knowledge scores and decisional conflict. Further studies are needed to determine the optimal way to implement these decision aids in clinical settings.
Combining Multiple Treatment: Advanced care planning - With trained oncology ranking, and list and discuss with assistance from a general adult US patient. Patients' utilities were assessed using a computerised self-adaptive conjoint analysis. After information about the different treatment options was provided, the participant was asked whether or not to have each treatment (e.g., chemotherapy, surgery, radiation therapy). The interview attempted to consider the value they attached to possible outcomes by rating factors and making pairwise comparison or not (participants randomized to receive decision analysis or not). Information about prenatal testing and/or diagnostic test(s) to undergo. Afterwards, the user can calculate the participant’s value of the chance of carrying a fetus and/or disease. The interview attempted to consider the value they attached to possible outcomes by rating factors and making pairwise comparison or not (participants randomized to receive decision analysis or not). Information about prenatal testing and/or diagnostic test(s) to undergo. Afterwards, the user can calculate the participant’s value of the chance of carrying a fetus and/or disease.

Controlled Trial: Reproductive health - Influence of a values discussion on health outcomes, and costs. A randomized controlled trial (RCT) was designed to explore the potential benefits of adding a values discussion to a decision-support guide. Participants were randomized to receive either a decision-support guide that included a values discussion (VCM) or a guide that did not include a values discussion (DA + VCM). The control group received usual care (control).patients fill out a questionnaire about their values and preferences regarding fertility preservation. The interview aimed to help patients clarify their values and preferences regarding fertility preservation. 3) No VCM (receiving usual care). 4) Education alone (video/leaflet only, 5) Decision analysis only; 6) VCM + Decision analysis (DA + VCM). 7) VCM- (DA with No VCM). 8) Decision analysis + VCM (information only vs VCM). Each participant was assigned a value to the different treatments depending on their personal preferences. No specific theory or strategy or any of the other decision-making approaches was considered to be an advantage or disadvantage compared to the decision analysis method. Our findings in this study indicated that the VCM group had a higher rate of decision regret or acceptability: the vast majority of participants reported that they would accept the decision-making process.
### Prioritization exercise

With assistance from a

#### Decision analysis

- Have a colostomy or die

- With binary

#### Effect of Adding a Values Clarification Risk Communication, Values

- Prioritization exercise

- Whether or not to be

### Sagittal Decision Aids for Prostate Cancer: A Randomized Trial

- Disease Prevention: A Randomized Exercise to a Decision Aid on Heart

### Avalos et al. 2013

Participants were then asked to make a decision about their medical situation and asked to render the values-decision did not align with their choice if their participants of the best fit for the disease, treatment choices, and specific wording of the

### Screening option and a fixed

A mathematical function is specified levels for each of these treatment attributes presented in the

### Sliders

- The slider positions.

- which received the

- Two groups: 1) the control arm which received

### Participants who

- completed a values

- Participants who

### Women and men ages 18-

- Unscrenned for prostate

### Pangume 2018

The effect was more likely to indicate

### Participants who

- never used hormone

### Decision aids: a systematic review and meta-analysis

- led to more people having

### Decision aids: a systematic review and meta-analysis

- showed that being in the

### Multivariable analyses

- showed that being in the

### Multivariable analyses

- showed that being in the

### Randomized to receive no

- which option fit best with

### Clinical characteristics

- led to more people having

### Multivariable analyses

- led to more people having

### Multivariable analyses

- led to more people having

### Data analysis

- led to more people having

### Randomized to receive no

- which option fit best with

### Women and men ages 18-

- Unscrenned for prostate

### Decision aids: a systematic review and meta-analysis

- led to more people having

### Women and men ages 18-

- Unscrenned for prostate

### Decision aids: a systematic review and meta-analysis

- led to more people having

### Women and men ages 18-

- Unscrenned for prostate

### Decision aids: a systematic review and meta-analysis

- led to more people having

### Women and men ages 18-

- Unscrenned for prostate

### Decision aids: a systematic review and meta-analysis

- led to more people having
### Table 1: Comparison of Strategies Across Six Studies

| Strategy | Type of Surgery | Cancer - Colon | Decisional Conflict | Values Congruent | Decisional Implications |
|----------|-----------------|----------------|--------------------|-----------------|------------------------|
| Strategy 6a | Unconstrained Dynamic Web Sliders | Randomized to receive no intervention or VCM | Compared to Strategy 6b, did result in reduced decisional conflict. | Offered no benefit in terms of values clarification. | A prototypical 'rating' of rating attributes. |
| Strategy 6b | Constrained Dynamic Web Sliders | Randomized to receive no intervention or VCM | Compared to Strategy 6a, resulted in increased values congruence compared to the control condition, but no benefit in terms of decisional implications. | Greater values congruence. | A prototypical 'rating' of rating attributes. |
| Strategy 6c | Constraint Dynamic Web Sliders | Randomized to receive no intervention or VCM | Compared to Strategy 6c, resulted in increased values congruence and decisional implications. | Reduced decisional conflict. | A prototypical 'rating' of rating attributes. |

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).
Online Appendix 3. Additional Results

Additional Meta-analytic Results

Figure S1. Values Congruence by Tradeoffs

| Study or Subgroup | VCM Events | Total | no VCM Events | Total | Weight | Risk Difference M-H, Random, 95% CI | Year |
|-------------------|------------|-------|---------------|-------|--------|--------------------------------------|------|
| Sheridan 2010     | 3          | 62    | 6             | 75    | 5.9%   | -0.03 [-0.11, 0.05]                  | 2010 |
| Jayadevappa 2015 | 17         | 343   | 21            | 350   | 18.7%  | -0.01 [-0.04, 0.02]                  | 2015 |
| Wittman 2015      | 85         | 205   | 110           | 202   | 4.4%   | -0.06 [-0.18, 0.06]                  | 2015 |
| Wittman 2020 6b   | 60         | 395   | 83            | 422   | 10.9%  | -0.05 [-0.10, 0.01]                  | 2020 |
| Wittman 2020 2a   | 43         | 239   | 59            | 217   | 6.5%   | -0.09 [-0.17, -0.02]                 | 2020 |
| Wittman 2020 2a + 2b | 26     | 237   | 41            | 217   | 8.4%   | -0.08 [-0.14, -0.01]                 | 2020 |
| Wittman 2020 6b + 6c | 69      | 413   | 93            | 422   | 11.6%  | -0.08 [-0.13, -0.03]                 | 2020 |
| Subtotal (95% CI) | 1894       | 423   | 1905          | 66.5% | -0.05 [-0.08, -0.02]                 |      |

Heterogeneity: Tau^2 = 0.00; Chi^2 = 10.41, df = 6 (P = 0.11); I^2 = 42%
Test for overall effect: Z = 3.49 (P = 0.0005)

2.2.2 Not at all

| Study or Subgroup | VCM Events | Total | no VCM Events | Total | Weight | Risk Difference M-H, Random, 95% CI | Year |
|-------------------|------------|-------|---------------|-------|--------|--------------------------------------|------|
| O'Conner 1999     | 35         | 101   | 33            | 100   | 2.6%   | 0.02 [-0.11, 0.15]                   | 1999 |
| Wittman 2020 4b   | 74         | 350   | 93            | 422   | 9.9%   | -0.01 [-0.07, 0.05]                  | 2020 |
| Wittman 2020 6a   | 85         | 263   | 93            | 422   | 9.8%   | 0.01 [-0.05, 0.07]                   | 2020 |
| Wittman 2020 6c   | 79         | 440   | 93            | 422   | 11.2%  | -0.04 [-0.08, 0.01]                  | 2020 |
| Subtotal (95% CI) | 1254       | 312   | 1366          | 33.5% | -0.01 [-0.04, 0.02]                 |      |

Total events: 373 312
Heterogeneity: Tau^2 = 0.00; Chi^2 = 3.06, df = 3 (P = 0.56); I^2 = 0%
Test for overall effect: Z = 0.75 (P = 0.45)

Total (95% CI): 3148 3271 100.0% -0.04 [-0.08, 0.00]

Total events: 368 735
Heterogeneity: Tau^2 = 0.00; Chi^2 = 13.88, df = 10 (P = 0.18); I^2 = 29%
Test for overall effect: Z = 3.42 (P = 0.0006)
Test for subgroups differences: Chi^2 = 3.47, df = 1 (P = 0.06); I^2 = 71.2%
### Figure S2. Values Congruence by Implementation/Presentation of Results

| Study or Subgroup | VCM Events | Total | no VCM Events | Total | Weight | Risk Difference M.H. Random, 95% CI | Year | Risk Difference M.H. Random, 95% CI |
|-------------------|------------|-------|---------------|-------|--------|-------------------------------------|------|-------------------------------------|
| **2.2.1 Presentation of results** |
| Wittman 2015      | 95         | 205   | 110           | 202   | 4.4%   | -0.06 [-0.10, 0.02]                | 2015 |                                    |
| Jayadevappa 2015  | 17         | 343   | 21            | 350   | 18.7%  | -0.01 [-0.04, 0.02]                | 2015 |                                    |
| Wittman 2020 6b + 6c | 56         | 413   | 93            | 422   | 11.6%  | -0.08 [-0.13, -0.03]               | 2020 |                                    |
| Wittman 2020 6c    | 79         | 443   | 93            | 422   | 11.2%  | -0.04 [-0.09, 0.01]                | 2020 |                                    |
| Wittman 2020 2a    | 43         | 239   | 59            | 217   | 6.5%   | -0.08 [-0.17, -0.02]               | 2020 |                                    |
| Wittman 2020 2a + 2b | 26         | 237   | 41            | 217   | 8.4%   | -0.08 [-0.14, -0.01]               | 2020 |                                    |
| **Subtotal (95% CI)** | 1877       | 1830  | 60.3%         |       |        | -0.06 [-0.09, -0.02]               |      |                                    |
| Total events      | 319        | 417   |               |       |        |                                    |      |                                    |
| Heterogeneity Tau² = 0.00; Chi² = 10.07, df = 5 (P = 0.07); I² = 50% |
| Test for overall effect: Z = 3.25 (P = 0.001) |

| **2.2.2 No presentation of results** |
| O'Connor 1999      | 35         | 101   | 33            | 100   | 2.6%   | 0.02 [-0.11, 0.15]                | 1999 |                                    |
| Sheridan 2010      | 3          | 62    | 6             | 75    | 5.9%   | -0.03 [-0.11, 0.05]               | 2010 |                                    |
| Wittman 2020 4b    | 74         | 350   | 93            | 422   | 9.9%   | -0.01 [-0.07, 0.05]               | 2020 |                                    |
| Wittman 2020 6a    | 85         | 363   | 93            | 422   | 9.0%   | 0.01 [0.05, 0.07]                 | 2020 |                                    |
| Wittman 2020 6b    | 68         | 395   | 93            | 422   | 10.9%  | -0.06 [-0.16, 0.01]               | 2020 |                                    |
| **Subtotal (95% CI)** | 1271       | 1441  | 39.2%         |       |        | -0.02 [-0.05, 0.01]               |      |                                    |
| Total events      | 285        | 310   |               |       |        |                                    |      |                                    |
| Heterogeneity Tau² = 0.00; Chi² = 2.01, df = 4 (P = 0.58); I² = 9% |
| Test for overall effect: Z = 1.10 (P = 0.27) |

| Total (95% CI)     | 3148       | 3271  | 100.0%        |       |        | 0.04 [-0.06, 0.02]               |      |                                    |
| Total events      | 584        | 735   |               |       |        |                                    |      |                                    |
| Heterogeneity Tau² = 0.00; Chi² = 13.88, df = 10 (P = 0.18); I² = 28% |
| Test for overall effect: Z = 3.42 (P = 0.0005) |
| Test for subarachnoid differences: Chi² = 2.92, df = 1 (P = 0.08); I² = 65.7% |
Figure S3. Values Congruence by Stated Use of a Theory/Framework

| Study or Subgroup | VCM Events | Total | no VCM Events | Total | Weight | Risk Difference M-H, Random, 95% CI | Year |
|-------------------|------------|-------|---------------|-------|--------|----------------------------------|------|
| **2.2.1 Stated theory** |
| Jayadevappa 2015  | 17         | 343   | 21            | 350   | 18.7%  | -0.01 [-0.04, 0.02]               | 2015 |
| Subtotal (95% CI) | 343        | 350   |               |       | 18.7%  | -0.01 [-0.04, 0.02]               |      |
| **Total events**  | 17         | 343   | 21            | 350   |        |                                  |      |
| Heterogeneity: Not applicable |
| Test for overall effect: Z = 0.00 (P = 0.99) |

| **2.2.2 No stated theory** |
| O'Connor 1999  | 35         | 104   | 6             | 75    | 6.9%   | -0.03 [-0.11, 0.05]               | 2010 |
| Sheridan 2010 | 3          | 62    | 6             | 75    | 6.9%   | -0.03 [-0.11, 0.05]               |      |
| Witteman 2015 | 95         | 205   | 110           | 202   | 4.4%   | -0.06 [-0.18, 0.02]               |      |
| Witteman 2020 2a | 43       | 239   | 59            | 217   | 6.5%   | -0.08 [-0.17, -0.02]              |      |
| Witteman 2020 2a + 2b | 28   | 237   | 41            | 217   | 8.4%   | -0.08 [-0.14, -0.01]              |      |
| Witteman 2020 4b | 74       | 350   | 93            | 422   | 9.9%   | -0.01 [-0.07, 0.05]               |      |
| Witteman 2020 6a | 85       | 363   | 83            | 422   | 9.8%   | 0.01 [0.05, 0.07]                 |      |
| Witteman 2020 6b | 68       | 395   | 93            | 422   | 10.9%  | -0.05 [-0.10, 0.01]               |      |
| Witteman 2020 6b + 6c | 69   | 413   | 93            | 422   | 11.6%  | -0.08 [-0.13, -0.03]              |      |
| Witteman 2020 6c | 79       | 440   | 93            | 422   | 11.2%  | -0.04 [-0.08, -0.01]              |      |
| Subtotal (95% CI) | 2805      | 2921  |               |       | 81.3%  | -0.04 [-0.07, -0.01]              |      |
| **Total events**  | 567       | 714   |               |       |        |                                  |      |
| Heterogeneity: Tau² = 0.00, Chi² = 10.83, df = 9 (P = 0.29), I² = 17% |
| Test for overall effect: Z = 3.79 (P = 0.0002) |
| **Total (95% CI)** | 3148      | 3271  | 100.0%        |       | -0.04 [-0.06, -0.02]              |      |
| **Total events**  | 584       | 735   |               |       |        |                                  |      |
| Heterogeneity: Tau² = 0.00, Chi² = 13.88, df = 10 (P = 0.18), I² = 28% |
| Test for overall effect: Z = 3.42 (P = 0.0006) |
| Test for subgroup differences: Chi² = 2.64, df = 1 (P = 0.10), I² = 62.1% |
Figure S4. Values Congruence by Real/Hypothetical Decision

| Study or Subgroup | VCM Events Total | no VCM Events Total | Risk Difference M-H, Random, 95% CI Year | Risk Difference M-H, Random, 95% CI |
|-------------------|------------------|---------------------|-----------------------------------------|-----------------------------------------|
| **2.2.1 Real decision** | | | | |
| O'Connor 1999     | 35 101           | 33 100              | 2.6% [0.11, 0.15] 1999                  |                                         |
| Jayadevagpa 2015  | 17 343           | 21 350              | 18.7% [-0.04, 0.02] 2015               |                                         |
| Wittman 2015      | 95 205           | 110 202             | 4.4% [0.18, 0.02] 2015                 |                                         |
| **Subtotal (95% Cl)** | 649 164         | 652 164             | **-0.02 [0.07, 0.03]**                  | **-0.02 [0.07, 0.03]**                  |
| **Total events**  | 147              | 164                 |                                         |                                         |
| Heterogeneity     | **Tau² = 0.00, Cb² = 2.67, df = 2 (P = 0.26), P² = 26%** | |                                         |                                         |
| Test for overall effect | Z = 0.86 (P = 0.39) | | | |

| **2.2.2 Hypothetical decision** | | | | |
| Sheridan 2010      | 3 62              | 6 75                | 5.9% [0.11, 0.05] 2010                 |                                         |
| Wittman 2020 4b    | 74 363            | 93 422              | 9.9% [0.07, 0.05] 2020                 |                                         |
| Wittman 2020 5a    | 85 363            | 93 422              | 9.8% [0.06, 0.07] 2020                 |                                         |
| Wittman 2020 6b    | 68 395            | 93 422              | 10.9% [0.10, 0.01] 2020                |                                         |
| Wittman 2020 6b + 6c | 55 413           | 93 422              | 11.6% [0.13, 0.03] 2020                |                                         |
| Wittman 2020 6c    | 75 440            | 93 422              | 11.2% [0.09, 0.01] 2020                |                                         |
| Wittman 2020 2a    | 43 239            | 59 217              | 6.5% [0.17, 0.02] 2020                 |                                         |
| Wittman 2020 2a + 2b | 75 237          | 41 217              | 8.4% [0.14, 0.01] 2020                 |                                         |
| **Subtotal (95% Cl)** | 2499 1240        | 2649 1210           | **-0.04 [0.07, 0.02]**                 | **-0.04 [0.07, 0.02]**                 |
| **Total events**  | 637              | 571                 |                                         |                                         |
| Heterogeneity     | **Tau² = 0.00, Cb² = 9.45, df = 7 (P = 0.22), P² = 26%** | |                                         |                                         |
| Test for overall effect | Z = 3.47 (P = 0.0005) | | | |
| **Total (95% Cl)** | 3148 1671        | 3217 1654           | **0.04 [0.06, 0.02]**                  | **0.04 [0.06, 0.02]**                  |
| **Total events**  | 584              | 755                 |                                         |                                         |
| Heterogeneity     | **Tau² = 0.00, Cb² = 13.88, df = 10 (P = 0.18), P² = 78%** | |                                         |                                         |
| Test for overall effect | Z = 3.42 (P = 0.0005) | | | |
| Test for subgroup differences: Cb² = 6.88, df = 1 (P = 0.41), P² = 0% | | | |
## Figure S5. Decisional Conflict by Tradeoffs

| Study or Subgroup | VCM Mean | VCM SD | Total | no VCM Mean | no VCM SD | Total | Weight | Std. Mean Difference IV, Random, 95% CI | Year |
|-------------------|---------|--------|-------|-------------|----------|-------|--------|----------------------------------------|------|
| Explicit tradeoffs | Montgomerie 2003 | 27.6 | 12.1 | 100 | 38.0 | 18.3 | 112 | 5.7% | -0.72 [-1.60, 0.14] | 2003 |
| Montgomerie 2007 | 23.8 | 16.1 | 100 | 27.8 | 14.8 | 201 | 6.0% | -0.28 [-0.48, -0.08] | 2007 |
| Fornell 2008 | 27.95 | 13.45 | 153 | 30 | 10.77 | 151 | 7.1% | -0.17 [-0.39, 0.06] | 2008 |
| Kuppermann 2014 | 12.9 | 14.1 | 357 | 13.8 | 15.6 | 353 | 9.7% | -0.06 [-0.21, 0.09] | 2014 |
| Garvelink 2014 | 41.6 | 9.5 | 66 | 44.1 | 12.3 | 68 | 4.4% | -0.23 [-0.57, 0.12] | 2014 |
| Shilk 2017 | 2.5 | 0.3 | 69 | 2.2 | 0.3 | 63 | 4.2% | 0.33 [-0.03, 0.69] | 2017 |
| Hopkins 2018 | 14.59 | 16.04 | 122 | 23.13 | 20.34 | 138 | 8.5% | -0.47 [-0.72, -0.22] | 2018 |
| Whiteman 2020 & | 36 | 27 | 413 | 42 | 27 | 432 | 10.1% | -0.22 [-0.36, -0.08] | 2020 |
| Whiteman 2020 63 | 39 | 27 | 395 | 42 | 27 | 423 | 10.0% | -0.11 [-0.25, 0.03] | 2020 |
| Whiteman 2020 64 | 35 | 25 | 350 | 42 | 27 | 442 | 9.9% | -0.25 [-0.40, -0.11] | 2020 |
| Subtotal (95% CI) | 2213 | 2348 | 75.7% | -0.22 [-0.34, -0.10] |

Heterogeneity: Tau² = 0.02; Chi² = 32.88, df = 9 (P = 0.0001); I² = 73%
Test for overall effect: Z = 3.67 (P = 0.0002)

Not at all

| Study or Subgroup | VCM Mean | VCM SD | Total | no VCM Mean | no VCM SD | Total | Weight | Std. Mean Difference IV, Random, 95% CI | Year |
|-------------------|---------|--------|-------|-------------|----------|-------|--------|----------------------------------------|------|
| Epstein 2018 | 0.09 | 2.25 | 61 | 9.43 | 2.23 | 61 | 4.2% | -0.17 [-0.52, 0.18] | 2018 |
| Whiteman 2020 63 | 36 | 26 | 440 | 42 | 27 | 423 | 10.2% | -0.23 [-0.36, -0.09] | 2020 |
| Whiteman 2020 64 | 35 | 27 | 363 | 42 | 27 | 422 | 9.9% | -0.26 [-0.40, -0.12] | 2020 |
| Subtotal (95% CI) | 864 | 905 | 24.3% | -0.24 [-0.33, -0.14] |

Heterogeneity: Tau² = 0.00; Chi² = 0.26, df = 2 (P = 0.88); I² = 0%
Test for overall effect: Z = 4.65 (P = 0.0000)

Total (95% CI) | 3077 | 3253 | 100.0% | -0.22 [-0.31, -0.13] |

Heterogeneity: Tau² = 0.02; Chi² = 33.47, df = 12 (P = 0.0008); I² = 64%
Test for overall effect: Z = 4.84 (P = 0.0000)
Test for subgroups differences: Chi² = 0.64, df = 1 (P = 0.85); I² = 0%
Figure S6. Decisional Conflict by Implications/Presentation of Results

| Study or Subgroup     | VCM Mean | VCM SD | VCM Total | no VCM Mean | no VCM SD | no VCM Total | Weight | IV, Random, 95% CI | Year |
|-----------------------|----------|--------|-----------|-------------|-----------|--------------|--------|--------------------|------|
| Montgomery 2003       | 27.6     | 12.1   | 100       | 38.3        | 18.3      | 112          | 5.7%   | -0.72 [-1.00, -0.44] | 2003 |
| Montgomery 2007       | 23.6     | 15.1   | 198       | 27.8        | 14.6      | 201          | 8.0%   | -0.28 [-0.48, -0.09] | 2007 |
| Rupprecht 2014        | 12.9     | 14.1   | 357       | 13.8        | 15.6      | 353          | 9.7%   | -0.06 [-0.21, 0.09]  | 2014 |
| Galvani 2014          | 41.6     | 4.5    | 66        | 12.3        | 44.1      | 68           | 4.4%   | -0.28 [-0.5, 0.22]   | 2014 |
| Shiff 2017            | 2.6      | 0.3    | 59        | 2.2         | 0.3       | 63           | 4.2%   | 0.33 [-0.63, 0.69]   | 2017 |
| Hopkin 2018           | 14.51    | 15.04  | 122       | 20.34       | 138       | 23.13        | 8.5%   | -0.47 [-0.72, -0.22] | 2019 |
| Wittman 2020 b        | 36       | 27     | 413       | 42          | 27        | 422          | 10.2%  | -0.23 [-0.36, -0.09] | 2020 |
| Wittman 2020 b & c    | 36       | 27     | 413       | 42          | 27        | 422          | 10.1%  | -0.22 [-0.36, -0.09] | 2020 |
| Subtotal (95% CI)     | 1755     | 1775   | 1938      | 10.0%       |           |              |        | -0.24 [-0.39, -0.09] |      |

Heterogeneity: $I^2 = 0.03$; $df = 7$ ($P = 0.99$); $Q = 77$%
Test for overall effect: $Z = 3.19$ ($P = 0.001$)

1.1.2 No presentation of results

| Study or Subgroup     | VCM Mean | VCM SD | VCM Total | no VCM Mean | no VCM SD | no VCM Total | Weight | IV, Random, 95% CI | Year |
|-----------------------|----------|--------|-----------|-------------|-----------|--------------|--------|--------------------|------|
| Frosh 2008            | 27.58    | 13.45  | 153       | 10.77       | 151       | 7.1%         | -0.17 [-0.39, 0.06] | 2008 |
| Epstein 2018          | 8.69     | 2.35   | 61        | 9.48        | 2.23      | 61           | 4.2%   | -0.17 [-0.52, 0.19] | 2018 |
| Wittman 2020 b        | 36       | 27     | 360       | 42          | 27        | 422          | 10.0%  | -0.11 [-0.26, 0.03] | 2020 |
| Wittman 2020 b & c    | 36       | 27     | 360       | 42          | 27        | 422          | 9.9%   | -0.28 [-0.40, -0.16] | 2020 |
| Subtotal (95% CI)     | 1322     | 1478   | 1180      | 41.3%       |           |              |        | -0.20 [-0.27, -0.13] |      |

Heterogeneity: $I^2 = 0.00$; $df = 4$ ($P = 0.60$); $P = 0$
Test for overall effect: $Z = 5.28$ ($P = 0.00001$)

Total (95% CI) 3077 3253 100.0% -0.22 [-0.31, -0.13]

Heterogeneity: $I^2 = 0.03$; $df = 12$ ($P = 0.0000$); $P = 64$
Test for overall effect: $Z = 4.94$ ($P = 0.00001$)
Test for subgroup differences: $I^2 = 0.24$; $df = 1$ ($P = 0.82$); $P = 0$
### Figure S7. Decisional Conflict by Stated Use of a Theory/Framework

| Study or Subgroup | VCM Mean | VCM SD | no VCM Mean | no VCM SD | Total Mean | Total SD | Weight | Std. Mean Difference IV, Random, 95% CI | Year |
|-------------------|----------|--------|-------------|----------|------------|----------|--------|---------------------------------------|------|
| Stated theory     |          |        |             |          |            |          |        |                                        |      |
| Montomery 2003    | 27.6     | 12.1   | 100         |          | 38.3       | 18.3     | 112    | -0.72 [-1.00, -0.44]                  | 2003 |
| Montgomery 2007   | 23.6     | 15.1   | 198         |          | 27.8       | 14.6     | 201    | -0.28 [-0.48, -0.08]                  | 2007 |
| Shik 2017         | 2.5      | 0.9    | 59          |          | 2.2        | 0.9      | 63     | 0.33 [0.03, 0.63]                     | 2017 |
| Epstein 2018      | 6.09     | 2.35   | 61          |          | 9.45       | 2.23     | 61     | -0.17 [-0.52, 0.19]                  | 2018 |
| Hopkins 2019      | 14.89    | 16.04  | 122         |          | 23.13      | 20.34    | 139    | -0.47 [-0.72, -0.22]                 | 2019 |
| Subtotal (95% CI) | 540      | 573    | 573         |          | 28.0%      |          | -0.28 [-0.57, 0.02]                  |      |

#### Heterogeneity
- Test for overall effect: Z = 1.85 (P = 0.06)

| No stated theory or framework | VCM Mean | VCM SD | no VCM Mean | no VCM SD | Total Mean | Total SD | Weight | Std. Mean Difference IV, Random, 95% CI | Year |
|-------------------------------|----------|--------|-------------|----------|------------|----------|--------|---------------------------------------|------|
| Frosh 2008                    | 27.95    | 13.45  | 153         |          | 30.10      | 17.77    | 151    | 7.1% [-0.39, 0.06]                    | 2008 |
| Gavrinix 2014                 | 41.8     | 9.5    | 66          |          | 44.1       | 12.3     | 68     | 4.4% [-0.57, 0.12]                   | 2014 |
| Kuppermann 2014              | 12.2     | 14.1   | 357         |          | 13.8       | 16.8     | 353    | 0.7% [-0.21, 0.00]                   | 2014 |
| Wittman 2020 6b               | 39       | 27     | 395         |          | 42         | 27       | 422    | 10.0% [-0.25, 0.00]                  | 2020 |
| Wittman 2020 6c               | 35       | 28     | 350         |          | 42         | 27       | 422    | 9.9% [-0.40, -0.11]                 | 2020 |
| Wittman 2020 6c               | 36       | 26     | 440         |          | 42         | 27       | 422    | 10.2% [-0.36, -0.02]                 | 2020 |
| Wittman 2020 6s               | 35       | 27     | 363         |          | 42         | 27       | 422    | 9.9% [-0.40, -0.12]                 | 2020 |
| Wittman 2020 6th + 8c         | 36       | 27     | 413         |          | 42         | 27       | 422    | 10.1% [-0.22, 0.06]                  | 2020 |
| Subtotal (95% CI)             | 1609     | 1680   | 1680        |          | 71.4%      |          | -0.18 [-0.24, -0.14]                 |      |

#### Heterogeneity
- Test for overall effect: Z = 8.55 (P < 0.001)

#### Total (95% CI)
- Test for overall effect: Z = 0.03 (P = 0.98)
- Test for subgroups difference: Chi² = 0.33, df = 1 (P = 0.57)
Figure S8. Decisional Conflict by Real/Hypothetical

| Study or Subgroup | VCM Mean | VCM SD | no VCM Mean | no VCM SD | Total Mean | Total SD | Total Weight | IV, Random, 95% CI Year |
|-------------------|----------|--------|-------------|----------|------------|----------|--------------|------------------------|
| 1.1.1 Real decision |          |        |             |          |            |          |              |                        |
| Montgomery 2003   | 27.6     | 12.1   | 100         | 38.0     | 18.3       | 112      | 5.7%         | -0.72 [-1.60, 0.14] 2009 |
| Montgomery 2007   | 25.6     | 15.1   | 100         | 27.8     | 14.0       | 201      | 6.0%         | -0.28 [-0.48, -0.08] 2007 |
| From 2008         | 27.95    | 13.45  | 153         | 30       | 10.77      | 151      | 7.1%         | -0.17 [-0.39, 0.06] 2008 |
| Kuppermann 2014   | 12.9     | 14.1   | 357         | 13.0     | 15.6       | 353      | 9.7%         | -0.06 [-0.21, 0.09] 2014 |
| Shirk 2017        | 2.5      | 0.9    | 59          | 2.2      | 0.9        | 63       | 4.2%         | 0.33 [0.03, 0.64] 2017 |
| Epstein 2018      | 0.00     | 2.35   | 61          | 9.4      | 2.23       | 61       | 4.2%         | -0.17 [-0.52, 0.20] 2018 |
| Subtotal (95% CI) | 928      | 941    | 39.0%       |          |            |          |              | -0.19 [-0.41, 0.03]    |

Heterogeneity: $I^2 = 0.08$; $Chi^2 = 25.78$, df = 5 ($P = 0.0021$); $I^2 = 01$
Test for overall effect: $Z = 1.68$ ($P = 0.10$)

1.1.2 Hypothetical decision

| Study or Subgroup | VCM Mean | VCM SD | no VCM Mean | no VCM SD | Total Mean | Total SD | Total Weight | IV, Random, 95% CI Year |
|-------------------|----------|--------|-------------|----------|------------|----------|--------------|------------------------|
| Garavello 2014    | 41.6     | 9.5    | 66          | 44.1     | 12.3       | 68       | 4.4%         | -0.23 [-0.57, 0.12] 2014 |
| Hopkin 2019      | 14.59    | 15.04  | 122         | 23.13    | 20.34      | 138      | 8.5%         | -0.47 [-0.72, -0.22] 2019 |
| White 2020 6b     | 39       | 27     | 395         | 42       | 27         | 422      | 10.0%        | -0.11 [-0.25, 0.03] 2020 |
| White 2020 4b     | 35       | 28     | 350         | 42       | 27         | 422      | 8.9%         | -0.26 [-0.40, -0.11] 2020 |
| White 2020 8b-8c  | 36       | 27     | 413         | 42       | 27         | 422      | 10.1%        | -0.22 [-0.38, -0.06] 2020 |
| White 2020 6a     | 35       | 27     | 363         | 42       | 27         | 422      | 9.9%         | -0.26 [-0.40, -0.12] 2020 |
| Subtotal (95% CI) | 2149     | 2312   | 61.0%       |          |            |          |              | -0.23 [-0.29, -0.17]    |

Heterogeneity: $I^2 = 0.00$; $Chi^2 = 0.85$, df = 6 ($P = 0.34$); $I^2 = 12$
Test for overall effect: $Z = 7.07$ ($P = 0.0001$)

Total (95% CI) | 3077 | 3253 | 100.0% | -0.22 [-0.31, -0.13] |

Heterogeneity: $I^2 = 0.02$; $Chi^2 = 33.47$, df = 12 ($P = 0.0008$); $I^2 = 64$
Test for overall effect: $Z = 4.94$ ($P = 0.0000$)

Test for subgroups differences: $Chi^2 = 0.13$, df = 1 ($P = 0.72$), $I^2 = 0$

Favours VCM | Favours no VCM
Figure S9. Decisional Conflict by Risk of Bias

| Study or Subgroup | VCM | no VCM | Std. Mean Difference | Year |
|------------------|-----|--------|----------------------|------|
|                  | Mean | SD     | Total               |      |
|                  | Mean | SD     | Total               |      |
|                  | Weight | IV, Random, 95% CI |      |
| Low risk of bias or some concerns |     |       |                     |      |
| Montgomery 2007 | 23.8 | 15.1 | 159     | 27.8 | 14.6 | 291 | -0.20 [-0.48, -0.01] | 2007 |
| Fresh 2008     | 27.95 | 13.45 | 153     | 30 | 13.77 | 151 | -0.17 [-0.35, 0.01] | 2008 |
| Gasolnik 2014 | 41.9 | 5.5 | 85     | 44.1 | 12.3 | 85 | -0.44 [-0.87, 0.01] | 2014 |
| Kuppermann 2014 | 12.5 | 14.1 | 357 | 13.9 | 15.6 | 353 | 0.11 [0.01, 0.21] | 2014 |
| Shirk 2017     | 1.9 | 0.9 | 59 | 2.2 | 0.9 | 63 | 0.03 [0.03, 0.03] | 2017 |
| Epstein 2018   | 9.08 | 2.35 | 81 | 9.48 | 2.23 | 81 | -0.17 [-0.32, 0.07] | 2018 |
| Hopfen 2018   | 14.6 | 16.4 | 122 | 23.1 | 20.3 | 128 | -0.47 [-0.72, -0.22] | 2018 |
| Wittmann 2020 | 35 | 27 |413 | 42 | 27 |422 | -0.22 [-0.36, -0.08] | 2018 |
| Wittmann 2020 | 36 | 26 | 440 | 42 | 27 | 422 | -0.23 [-0.36, -0.09] | 2020 |
| Wittmann 2020 | 35 | 26 | 350 | 42 | 27 | 422 | -0.26 [-0.40, -0.11] | 2020 |
| Wittmann 2020 | 35 | 26 | 365 | 42 | 27 | 422 | -0.26 [-0.40, -0.11] | 2020 |
| Wittmann 2020 | 35 | 26 | 395 | 42 | 27 | 422 | -0.11 [-0.26, 0.04] | 2020 |
| Subtotal (95% CI) | 2977 | 3140 | 89.7% | -0.20 [-0.27, -0.12] |      |

Heterogeneity: Tau² = 0.00, Chi² = 20.46, df = 11 (P = 0.04), I² = 46%
Test for overall effect: Z = 5.29 (P = 0.00001)

High risk of bias

| Study or Subgroup | VCM | no VCM | Std. Mean Difference | Year |
|------------------|-----|--------|----------------------|------|
|                  | Mean | SD     | Total               |      |
|                  | Mean | SD     | Total               |      |
|                  | Weight | IV, Random, 95% CI |      |
| Montgomery 2003 | 27.8 | 12.1 | 150 | 38.9 | 19.3 | 112 | -0.57 [-0.99, -0.15] | 2003 |
| Gasolnik 2014 | 43.8 | 14.2 | 70 | 40.9 | 11.6 | 70 | 0.21 [0.02, 0.40] | 2014 |
| Subtotal (95% CI) | 170 | 182 | 10.3% | -0.26 [-1.17, 0.65] |      |

Heterogeneity: Tau² = 0.40, Chi² = 17.56, df = 1 (P < 0.0001), I² = 94%
Test for overall effect: Z = 0.35 (P = 0.73)

Total (95% CI) | 2147 | 3322 | 100.0% | -0.20 [-0.29, -0.11] |      |

Heterogeneity: Tau² = 0.02, Chi² = 39.60, df = 13 (P = 0.002), I² = 67%
Test for overall effect: Z = 4.33 (P < 0.0001)
Test for subgroups: Z = 0.02, df = 1 (P = 0.89), I² = 0%
Head-to-Head Evaluations of Values Clarification Methods

Feldman-Stewart et al. (2006) found no difference across all three groups (information only; values clarification method without a summary bar, i.e. rating scales; values clarification with a summary bar, i.e. multicriteria decision analysis) in terms of the attributes participants identified as important to their decisions nor in how difficult it was to make the decision. When trial participants were unblinded at the end of the study and shown all three options, all of them ranked the bars with the summary option (multicriteria decision analysis) as the most helpful.

Pignone et al. (2012) found that a discrete choice experiment produced somewhat different patterns of attribute importance compared to ranking and rating. Agreement between the most important attribute derived from the values clarification method and the most important attribute as reported by participants in the questionnaire was slightly higher in the ranking and rating arm than the discrete choice experiment arm. The authors found no difference between study arms in terms of values clarity, intent to be screened and unlabelled screening test preference.

Pignone et al. (2013) found that different values clarification methods produced differences in attribute importance and screening test preference. Participants who received the rating and ranking test were more likely to report the chance of dying from prostate cancer as the most important attribute compared to the balance sheet and discrete choice experiment groups. Those who received the balance sheet were more likely to prefer the unlabelled PSA-like test option compared to the two other groups. Participants who received the discrete choice experiment were somewhat less likely to select reduction of mortality as the most important attribute, and were least likely to select the PSA-like option on the unlabelled preference question. There was no difference across groups in intent to be screened (labelled PSA test option) nor on values clarity.

Brenner et al. (2014) found that different values clarification methods produced different results in terms of individuals’ most important screening test attributes. Specifically, respondents who received the rating and ranking exercise, compared to a discrete choice experiment or a balance sheet (i.e., implicit values clarification method), were the most likely to choose risk reduction as the most important attribute. They found no differences in terms of test preferences, values clarity, nor intention to be screened.

Witteman et al. (2020) found that overall, methods using mathematical models (e.g., decision analysis, allocation of points) were more promising than other methods (e.g., pros and cons, rating scales) for encouraging values-congruent decisions. All methods encouraged lower decisional conflict when this was assessed.
| Article                  | Random Sequence Generation (Selection Bias) | Allocation Concealment (Selection Bias) | Blinding of Participants and Personnel (Performance Bias) | Blinding of Outcome Assessment (Detection Bias) | Incomplete Outcome Data (Attrition Data) | Selective Reporting (Reporting Bias) | Other Bias |
|-------------------------|-------------------------------------------|-----------------------------------------|----------------------------------------------------------|------------------------------------------------|----------------------------------------|-------------------------------------|------------|
| Abhyankar 2010          | low                                       | unclear                                 | high                                                     | unclear                                        | low                                    | low                                 | low        |
| Au 2012                 | low                                       | low                                     | unclear                                                  | unclear                                        | low                                    | low                                 | unclear    |
| Bekker 2004             | low                                       | low                                     | unclear                                                  | unclear                                        | low                                    | low                                 | low        |
| Brenner 2014            | low                                       | low                                     | low                                                      | unclear                                        | low                                    | low                                 | low        |
| Clancy 1988             | low                                       | unclear                                 | unclear                                                  | low                                            | unclear                                | high                                | low        |
| deAchaval 2012          | low                                       | low                                     | low                                                      | low                                            | low                                    | low                                 | low        |
| Epstein 2018            | low                                       | unclear                                 | unclear                                                  | high                                            | unclear                                | low                                 | unclear    |
| Feldman-Stewart 2006    | unclear                                   | unclear                                 | unclear                                                  | unclear                                        | unclear                                | unclear                             | low        |
| Feldman-Stewart 2012    | low                                       | low                                     | low                                                      | low                                            | unclear                                | low                                 | low        |
| Fraenken 2007           | low                                       | low                                     | high                                                     | unclear                                        | low                                    | low                                 | low        |
| Frosch 2008             | low                                       | low                                     | unclear                                                  | low                                            | unclear                                | low                                 | low        |
| Garvelink 2014 (1)      | low                                       | low                                     | low                                                      | unclear                                        | low                                    | high                                |            |
| Garvelink 2014 (2)      | low                                       | low                                     | low                                                      | low                                            | low                                    | low                                 | unclear    |
| Hess 2015               | low                                       | low                                     | unclear                                                  | unclear                                        | low                                    | unknown                             |            |
| Hopkin 2019             | low                                       | low                                     | low                                                      | low                                            | unclear                                | low                                 | unclear    |
| Hutyra 2019             | low                                       | unclear                                 | low                                                      | low                                            | unclear                                | low                                 | unclear    |
| Jayadevappa 2019        | low                                       | low                                     | unclear                                                  | low                                            | unclear                                | low                                 | unclear    |
| Kennedy 2002            | low                                       | high                                    | unclear                                                  | low                                            | unclear                                | low                                 | unclear    |
| Kuppermann 2014         | low                                       | low                                     | low                                                      | low                                            | low                                    | unclear                             |            |
| Lehrman 1997            | unclear                                   | unclear                                 | high                                                     | unclear                                        | low                                    | low                                 | low        |
| Matheis-Kraft 1997      | unclear                                   | unclear                                 | unclear                                                  | unclear                                        | low                                    | unclear                             |            |
| Montgomery 2007         | low                                       | low                                     | low                                                      | low                                            | low                                    | low                                 | unclear    |
| Montgomery 2003         | low                                       | low                                     | high                                                     | low                                            | low                                    | unclear                             | low        |
| Reference       | Quality | Bias | Compliance | Conf | Methodology | Sample Size | Follow-Up |
|-----------------|---------|------|------------|------|-------------|-------------|-----------|
| Myers 2003      | unclear| unclear| unclear | unclear | low         | unclear     | unclear   |
| Myers 2005      | unclear| unclear| unclear | unclear | low         | low         | unclear   |
| O'Connor 1999   | low    | low  | unclear   | low   | low         | unclear     | low       |
| Paquin 2018     | low    | low  | unclear   | low   | low         | unclear     | low       |
| Peinado 2020    | unclear| low  | unclear   | unclear | low         | low         | low       |
| Pignone 2012    | unclear| unclear| low      | unclear | low         | unclear     | low       |
| Pignone 2013    | low    | low  | low       | unclear | low         | unclear     | low       |
| Sheridan 2010   | low    | low  | low       | low   | low         | low         | low       |
| Shirk 2017      | unclear| unclear| unclear | unclear | low         | low         | unclear   |
| Witteman 2015   | low    | low  | unclear   | unclear | low         | unclear     | unclear   |
| Witteman 2020   | unclear| low  | unclear   | unclear | low         | low         | low       |
| Section/topic | # | Checklist item                                                                                                                                                                                                 | Reported on page # |
|--------------|---|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| TITLE        |   | Title                                                                                                                                                                                                       | Title page        |
|              | 1 | Identify the report as a systematic review, meta-analysis, or both.                                                                                                                                           |                   |
| ABSTRACT     |   | Structured summary                                                                                                                                                                                             | 2                 |
|              | 2 | Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number. |                   |
| INTRODUCTION |   | Rationale                                                                                                                                                                                                   | 3-5               |
|              | 3 | Describe the rationale for the review in the context of what is already known.                                                                                                                               |                   |
|              | 4 | Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).                                                       | 3-5               |
| METHODS      |   | Protocol and registration                                                                                                                                                                                     | na                |
|              | 5 | Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.                                         |                   |
|              | 6 | Eligibility criteria                                                                                                                                                                                           | 6                 |
|              | 7 | Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.         |                   |
|              | 8 | Information sources                                                                                                                                                                                             | 6-7               |
|              | 9 | Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.                                 |                   |
|              | 10| Search                                                                                                                                                                                                       | Appendix 1       |
|              | 11| Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.                                                                            |                   |
|              | 12| Study selection                                                                                                                                                                                                 | 7                 |
|              | 13| State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).                                                      |                   |
|              | 14| Data collection process                                                                                                                                                                                         | 7                 |
|              |   | Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.                                     |                   |
|              | 15| Data items                                                                                                                                                                                                  | 7-8               |
|              |   | List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.                                                                         |                   |
|              | 16| Risk of bias in individual studies                                                                                                                                                                              | 8                 |
|              |   | Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis. |                   |
|              | 17| Summary measures                                                                                                                                                                                               | 8-9               |
|              |   | State the principal summary measures (e.g., risk ratio, difference in means).                                                                                                                                  |                   |
|              | 18| Synthesis of results                                                                                                                                                                                           | 8-9               |
|              |   | Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$) for each meta-analysis.                                                     |                   |
| Section/topic                  | #  | Checklist item                                                                                                                                                                                                 | Reported on page # |
|-------------------------------|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| Risk of bias across studies  | 15 | Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).                                                                     | 8                 |
| Additional analyses           | 16 | Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.                                                                        | 8-9               |
| RESULTS                       |    |                                                                                                                                                                                                                     |                   |
| Study selection               | 17 | Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.                                                      | 9                 |
| Study characteristics         | 18 | For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.                                                                           | 10-14             |
| Risk of bias within studies   | 19 | Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).                                                                                                          | 14                 |
| Results of individual studies | 20 | For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot. | 14-17             |
| Synthesis of results          | 21 | Present results of each meta-analysis done, including confidence intervals and measures of consistency.                                                                                                            | 14-17             |
| Risk of bias across studies   | 22 | Present results of any assessment of risk of bias across studies (see Item 15).                                                                                                                                     | Appendix 3        |
| Additional analysis           | 23 | Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).                                                                                                  | 14-17             |
| DISCUSSION                    |    |                                                                                                                                                                                                                     |                   |
| Summary of evidence           | 24 | Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).                                           | 17-18             |
| Limitations                   | 25 | Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).                                                     | 18-19             |
| Conclusions                   | 26 | Provide a general interpretation of the results in the context of other evidence, and implications for future research.                                                                                           | 19                 |
| FUNDING                       |    |                                                                                                                                                                                                                     |                   |
| Funding                       | 27 | Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.                                                                     | 20                 |

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097