



This is a repository copy of *Taking patient heterogeneity and preferences into account in health technology assessments*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/124666/>

Version: Accepted Version

Article:

Kievit, W., Tummers, M., van Hoorn, R. et al. (7 more authors) (2017) Taking patient heterogeneity and preferences into account in health technology assessments. International Journal of Technology Assessment in Health Care. ISSN 0266-4623

<https://doi.org/10.1017/S0266462317000885>

This article has been published in a revised form in International Journal of Technology Assessment in Health Care [<https://doi.org/10.1017/S0266462317000885>]. This version is free to view and download for private research and study only. Not for re-distribution, re-sale or use in derivative works. © Cambridge University Press 2017.

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Taking patient heterogeneity and preferences into account in Health Technology Assessments

Short title: patient heterogeneity and preferences in HTA

Authors: Wietske Kievit¹, Marcia Tummers¹, Ralph van Hoorn¹, Andrew Booth², Kati Mozygemba³, Pietro Refolo⁴, Dario Sacchini⁴, Lisa Pfadenhauer⁵, Ansgar Gerhardus³, Gert Jan Van der Wilt⁶

¹Radboud Institute for Health Sciences, Radboud university medical center, Nijmegen, the Netherlands

²Health Economics and Decision Science (HEDS), School of Health and Related Research (ScHARR), University of Sheffield Regent Court, Sheffield, United Kingdom

³Institute for Public Health and Nursing Research (IPP), University of Bremen/Health Sciences Bremen, Bremen, Germany

⁴Institute of Bioethics, A. Gemelli School of Medicine, Università Cattolica del Sacro Cuore, Rome, Italy

⁵Institute for Biometrics, Epidemiology and Medical Informatics, Ludwig-Maximilians-Universität München, Munchen, Germany

⁶Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands

Corresponding author:

Wietske Kievit PhD

Radboud Institute for Health Sciences

Department for Health Evidence

Radboud university medical center

Postbus 9101, 6500 HB Nijmegen, The Netherlands

Tel: +31243610592;

Email: wietske.kievit@radboudumc.nl

Funding: this project was funded by the European Union under the Seventh Framework Programme (FP7 -Health-2012-Innovation) under grant agreement number 306141

Abstract

Introduction: The INTEGRATE-HTA project provided methodology to evaluate complex technologies. This paper provides guidance on how to retrieve and critically appraise available evidence on moderators and predictors of treatment effects and on patient preferences for treatment outcomes as a source of complexity.

Methods: search filters for PubMed were developed by hand-searching a large volume of articles reporting on relevant aspects. Search terms were retrieved from selected papers and algorithmically combined to find the optimal combination of search terms. For the development of the appraisal checklists literature was searched in PubMed and Google Scholar together with citation chasing. For the Checklist for the Appraisal of Moderators and Predictors (CHAMP) a Delphi procedure was used to value a set of eligible appraisal criteria retrieved from the literature.

Results: search filters were developed optimized for different accuracy measures. The final version of CHAMP consists of a 17 questions covering the design, analysis, results and transferability of results of moderator and predictor analysis. The final checklist for appraisal of literature on patient preferences for treatment outcomes consist of six questions meant to help the user to indentify relevant quality issues together with a guidance towards existing tools concerning the appraisal of specific preference elicitation methods.

Conclusion: Incorporating knowledge on subgroups for whom a specific treatment will produce more benefit holds the promise of better targeting and, ultimately, enhancing overall effectiveness and efficiency of healthcare technology. Finally, incorporating information on preferences for treatment outcomes will foster HTA that addresses outcomes that are important to patients.

Keywords: INTEGRATE-HTA, moderators, predictors, patient preferences, complex interventions, health technology assessment

Introduction

Traditionally, Health Technology Assessment (HTA) focuses on effectiveness and cost-effectiveness while other domains as part of the original definition of HTA such as ethical, legal and social aspects have been neglected. Within the INTEGRATE-HTA project we advocate for integration of social cultural, ethical, legal, organizational issues with effectiveness and cost-effectiveness (1). Social cultural, ethical and legal aspects present within the context may interact with the intervention which might make an intervention complex. If an HTA process fails to take into account the complexity of an intervention then it could result in misleading conclusions (1).

Another source of complexity, studied in work package 4 of the INTEGRATE-HTA project, is heterogeneity in patient characteristics and preferences for treatment outcomes. Depending on their characteristics patients may respond differently to a specific treatment, both in terms of beneficial or adverse effects. Moreover, patients may differ in how they value particular treatment outcomes. For instance, in the context of treatment of patients with epilepsy, a specific drug may be known to lead, on average, to slightly superior control of seizures, improved mood, but also weight gain. Even though this superiority may be true on average, seizure control will be better achieved in some patients using other drugs. Also, weight gain, if it occurs, may be less of a problem to some patients when compared with others.

This example illustrates the need for information about moderators and predictors of treatment effects as well as patient preferences for specific treatment outcomes. First, information about moderators or predictors can be used to guide the search for subgroup analyses. Furthermore, the effectiveness of an intervention under assessment could be enhanced by targeting the treatment to a group that is most likely to benefit and thus guide implementation of a new intervention. Information about patient preferences for treatment outcome may be used to guide the effectiveness evaluation so as to ensure that this preferred outcome becomes one of the most important outcomes in the evaluation. Furthermore, in the decision making process following an HTA, one could give greater weight to results on outcomes that matter to patients most.

The objective of work package 4 of the INTEGRATE-HTA project was to develop tools to efficiently retrieve and critical appraise available evidence on (i) moderators and predictors of treatment effects as well as on (ii) patient preferences for treatment outcomes, using search filters and critical appraisal tools. This paper takes the perspective of HTA researchers who wish to use the best available evidence to develop recommendations about how, and for whom, healthcare technologies may be optimally targeted. This paper summarizes previously published guidance (2) and offers insight into how the tools can be used by HTA researchers or agencies.

Moderators and predictors of treatment effects

Moderators are variables which influence the strength of a relation between two other variables, for instance those for a treatment and an effect (3-5). Age and gender are common moderators. The term 'moderator' is similar to the epidemiological term 'effect modifier' and the statistical term 'interaction' (6). In practice it could mean that a treatment could work for men but not equally well for women. *Predictors* are characteristics or variables that influence the outcome independent of the treatment. The effect would be the same if a different treatment was applied. For instance gender is a predictor for mortality but when treatment had equal effects in men and women, gender is a predictor and not a moderator. Moderators can be investigated using subgroup analyses or regression analysis using interaction terms. Predictors can be investigated by looking at measures of association as in regression analyses (7-9).

Search filters

Efficiently identifying and using information from the literature concerning moderators or predictors of treatment effect requires an appropriate search strategy. As an alternative for hand searching the literature, a well defined search strategy is more efficient in retrieving relevant articles and allows for replication of results by its transparency. PubMed Clinical Query (PCQ) filters have been developed to find literature on prognosis, treatment or clinical prediction guides. These filters produce results containing articles concerning diagnosis or disease stage, study design/methodology, clinical prediction (i.e. prognosis, independent of treatment), outcome measures (including patient reported outcomes and quality of life) or treatment effects in general. A search filter specifically aimed at retrieving studies reporting on moderators of treatment outcome (i.e. variables that influence how well or likely a patient responds to a treatment) could have added value to existing filters.

Such a search filter was created by first collecting relevant articles on moderators and predictors of treatment effects. All articles published in the year 2011 in specific journals in the field of rheumatoid arthritis and general medicine were searched by hand and any of these reporting moderators or predictors of treatment effects were selected. Subsequently, search terms identified from these papers were algorithmically combined to derive the optimal combination of search terms for finding articles on moderators and predictors of treatment effects. The applied methods followed accepted good practice in search filter creation (10, 11). More details on the development on the search filters can be found in the Guidance for the assessment of treatment moderators and patients' preferences (2).

This methodology resulted in four sets of search filters, which were tested on the same hand searched set of papers. The results are listed in table 1. Each set contains the top three search filters optimized for the respective performance measure. If the purpose of an HTA is to be exhaustive, the filter optimized for sensitivity will be most appropriate. However it will probably return a relatively large proportion of papers of low relevance. Search filters with high accuracy, specificity or low number of papers needed to screen (Number Needed to Read (NNR) (12) will return fewer irrelevant papers at the expense of missing potentially important information. Clearly, the choice of which strategy to use depends on the goal of the systematic review, the numbers of usable retrieved papers and the amount of time that the user is willing and able to invest. The chosen search filter needs to be combined with a disease specific search filter relevant to the field of interest. We do not recommend to combine the search filters with any limits on publication type, since moderators and predictors are reported in epidemiological studies, randomised controlled trials (RCTs) as well as meta-analysis (7-9).

Critical appraisal tool

Once relevant literature has been selected, the next step is to assess it for quality. Two critical appraisal tools for prediction and moderation of treatment effect, based on Randomised Controlled Trials (RCTs), have previously been published (13, 14). However, relevant and valid moderator or predictor information is also found in studies other than RCTs. Also, such information may be derived from a body of evidence, as represented within a systematic review, rather than solely from single studies. Therefore, within work package 4 of the INTEGRATE-HTA project, we developed a more flexible critical appraisal tool suitable for all alternatives of moderator/ predictor (or subgroup) analyses.

For the development of the tool, we conducted a preliminary literature review of aspects important for validity of moderators and predictors of treatment effect. This review was based on searches in PubMed and Google Scholar, citation chasing, author searches, related articles and consultation with experts. As the aim was not to itemise every single aspect of validity, but simply to identify a diverse range of indicative factors, we did not aim for comprehensive coverage of the literature. Forty-nine appraisal criteria were identified in the literature. Subsequently, a Delphi procedure with three rounds was used, following the Research and Development (RAND) Appropriateness Method guidelines (15), to augment and then value a set of appraisal criteria retrieved from the literature. Fourteen experts from (bio)statistics, epidemiology and other associated fields participated in the Delphi procedure. Based on these results, a final selection of criteria was included in a test version of the appraisal checklist. Following internal testing and

external feedback the test version was amended to create a final version. We would like to refer to the Guidance for the assessment of treatment moderators and patients' preferences (2) for more details about the development of the tool and the results of the difference Delphi rounds and testing phases. Please note that work on the checklist was completed after the guidance was published. We are currently working on an update of the guidance.

The final version of the CHecklist for the Appraisal of Moderators and Predictors (CHAMP) consists of a 17 questions (listed in table 3), each completed as 'yes', 'no', 'don't know' or 'not applicable'. Those 17 items cover the design (e.g. *a priori* plausibility), analysis (e.g. use of interaction tests) and results (e.g. complete reporting) of moderator and predictor analysis, together with the transferability of the results. Using the checklist should help to arrive at a transparent and uniform overall judgement of the quality of a moderator or predictor analysis. CHAMP can be used to determine whether evidence is sufficient to warrant subgroup analyses in meta-analyses; to systematically value and describe evidence in systematic reviews; to design prediction models, and to facilitate individualised healthcare. CHAMP is designed to be used in conjunction with a quality tool such as the Cochrane risk of bias tool (16), to judge the overall quality of the study.

Patient preferences for treatment outcomes

The value of a specific technology for a defined individual does not only depend on moderators and predictors but also on their personal preferences. The importance of incorporating patients' preferences in medical decision making is increasingly recognized. The importance of patient preferences for treatment outcomes can be illustrated by the example of an HTA of the paediatric cochlear implant. Whereas the literature mainly reported outcomes for hearing and speech, the deaf community was at least equally interested in social and emotional development outcomes (17). In a further example, patients with chronic kidney disease differed regarding the weight they assigned to various hemodialysis-related outcomes when compared to the views of nephrologists and HTA authors (18). Both examples illustrate that interventions may be considered superior in aspects deemed important to medical professionals or decision makers but not to patients. The value of interventions should, therefore, also be established from the viewpoint of the target population, i.e. the patients.

Patients' preferences are usually described as a preference for one treatment or another. These preferences are difficult to generalize as they are highly context-dependent (19). Therefore, it is more relevant to retrieve information on treatment outcomes which might explain such

preferences, e.g. risks on adverse events, or specific outcomes such as functional status. Searching for information on preferences for treatment outcomes in the medical literature, for instance using PubMed, can be time-consuming (20, 21) and may be problematic since patient preferences are elicited in many ways (21, 22). Heterogeneity in methods used and reporting styles makes it more difficult to retrieve relevant literature (23). Therefore, we developed a search filter, similar to PubMed's Clinical Queries, with high performance in retrieving scientific papers that report empirical evidence on patients' preferences for treatment outcomes (24).

Search filters

Development of search filters for patient preferences followed a similar process to that for the filters for finding moderators and predictors: [1] a comprehensive set of search terms and combinations of terms was constructed and [2] the results of these combinations of terms were tested in a set of relevant papers. This methodology resulted in a set of search filters either optimized for sensitivity, specificity, accuracy or NNR as shown in Table 2 (24).

Testing revealed that papers on patient preferences in general and for treatment outcomes, specifically, are a needle in the medical literature haystack. Only 22 of all 8238 hand-searched articles (0.27%) reported empirical evidence on patients' preferences for treatment outcomes. We identified three possible reasons for this finding: i) there is little research performed on this subject; ii) the research is inadequately reported and cannot be retrieved at a title or abstract level; or iii) the journals we had carefully preselected do not commonly publish papers on preferences. Based on this finding we recommend starting with the sensitivity-optimised filters. When the initial set of retrieved literature seems unmanageably large, then a specificity-optimised filter can be used.

Inevitably, the performance of the search filters presented in this paper reflects the terminology used by researchers to publish their work on patients' preferences for treatment outcomes in 2011. Any changes in terminology over time affects the performance of search filters. For this reason, ongoing update of the performance of these search filters on a periodic basis is warranted.

Critical appraisal checklist

The aim of the appraisal checklist is to determine whether a study reporting on patient preferences for treatment outcomes has (i) been executed rigorously and (ii) whether the findings are relevant to the research questions of the HTA. Given the diversity of methods to elicit preferences, quantitative as well as qualitative, it did not seem feasible to develop *one* generic tool including items relevant to *all* study designs. Therefore, we mapped the methods currently being used to elicit patient

preferences for treatment outcomes and then searched for existing guidance or tools to appraise these methods.

To explore methodologies most commonly used to elicit patient preferences for treatment outcome, we analysed the papers identified during development of the search strategy, as well as expert opinion, and conducted additional PubMed and Google Scholar searches. A separate search was performed, for each method found, to identify appraisal criteria specific to that method. These searches combined method-related search terms with appraisal related search terms, such as 'appraisal' or 'quality'. The search identified various studies that detail quality criteria of potential value when appraising studies on patient preferences for treatment outcomes.

Despite the large variety of methods available to elicit patient preferences (25), we identified considerable overlap in how data are collected or interpreted across methods. Grouping of appraisal criteria was performed primarily on a conceptual, not a methodological, basis. Following creation of a test version, the tool was tested in a case study and revised following user feedback.

The final checklist consists of six questions (listed in table 4), each completed as 'yes', 'no', 'don't know' or 'not applicable'. The checklist, including an in-depth explanation of the rationale behind each question, is presented in the supplementary file. By answering the individual items, users should be able to identify relevant quality issues. The items in the checklist can be considered as a set of key quality indicators: the more these criteria are met, the greater the likelihood that a study was adequately conducted. For appraisal of specific aspects, or to determine the appropriateness of the method, will require in-depth knowledge of the specific methods used.

Application of the tools within INTEGRATE-HTA

In terms of the INTEGRATE-HTA process (26), information about moderators and predictors for treatment effects as well as the patient preferences for treatment outcome are best used as input for a logic model (27). A logic model can be used to conceptualize the complexity of a technology by making a graphical description of a system within which the technologies operate, its elements and any relationships within the system. On the one hand, information on preferences and potential moderators or predictors can be use to guide the effectiveness assessment. On the other hand, social cultural, ethical, or legal issues may determine or moderate either the preferences for treatment outcomes or treatment effects itself. This information may be looped back into the INTEGRATE-HTA process model (26) and into the logic model, making it a more comprehensive, iterative and integrated process.

This process is illustrated by the HTA on reinforced models of palliative care which served as a case study in the INTEGRATE-HTA project (28). The developed search filters and appraisal tools

were tested within this case study. According to the evidence retrieved, many patients receiving palliative care expressed a strong preference to die at home, in their familiar surroundings, an outcome not to be neglected when comparing different modalities of palliative care. Furthermore, the logic model should incorporate factors that may affect the likelihood of patients dying in their own familiar surroundings, as for example the presence of an informal caregiver.

In addition, we found papers reporting on factors that influence preferences for treatment outcomes. For example, in the North-American setting, blacks and Hispanics are more likely to want to spend their last days in a hospital, and more likely to want life-prolonging drugs, when compared with Caucasians (29). This would mean that a home-based palliative care intervention does not *per se* fit the needs of patients from this ethnic background and will probably not result in the best outcomes for this group.

Incorporating such information within an HTA may lead to a more targeted indication for particular services, as opposed to a “one size fits all” approach. However, numerous, diverse factors could influence clinical decision making, especially when a complex intervention is involved. Each additional factor incorporated within the decision making process adds to complexity and may incur additional costs. Especially when determining a genetic profile, specific biomarkers or laboratory tests will add to the overall costs of the technology. Prior to implementation decision-makers will need to consider whether the extra costs of a more individualized approach will outweigh the benefits. However, use of information on treatment moderating and patient preferences in clinical decision-making holds the potential to improve the quality of care, efficiency of care and saving costs (30, 31).

The proposed approach of INTEGRATE-HTA and in particular the results of work package 4 are in contrast with the traditional HTA where the cost-effectiveness analysis plays central role. If a cost-effectiveness analysis is done from a societal perspective, which is preferred by most guideline of CE analyses, health states are valued from the perspective of the general population. Indeed from a societal perspective, accounting for fairness, societal values are preferred but from an individual perspective of the patient all kind of other outcomes could be preferred. Including patient preferences in HTA can increase public acceptance of health policy, increase transparency and legitimacy by involving stakeholders, and is therefore essential to good HTA practice (32). Therefore, we highly recommend that information about treatment moderation and patient preferences is incorporated within an HTA in order to target populations that will benefit the most and to assess value of outcomes that are prioritized by patients.

Conclusion

This INTEGRATE-HTA work package 4 resulted in different tools to retrieve and critically appraise literature on moderators and predictors of treatment effects or patient preferences for treatment outcomes. Using the tools enables HTA researchers to retrieve information on subgroups for whom a specific treatment will produce more benefit. Incorporating this knowledge in the HTA process holds the promise of better targeting and, ultimately, enhancing overall effectiveness and efficiency of healthcare technology. Finally, incorporating information on preferences for treatment outcomes will foster HTA that addresses outcomes that are important to patients.

References

1. Gerhardus A. Integrated health technology assessment for evaluating complex technologies (INTEGRATE-HTA): An introduction to the guidances. 2016. Available via: <http://www.integrate-hta.eu/downloads/> (accessed on 20st of July 2017)
2. Van Hoorn RT, M.; Kievit, W.; Van der Wilt, G.J. Guidance for the assessment of treatment moderation and patients' preferences. 2016. Available via: <http://www.integrate-hta.eu/downloads/> (accessed on 20st of July 2017)
3. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*. 2002;59:877-83.
4. Nicholson RA, Hursey KG, Nash JM. Moderators and mediators of behavioral treatment for headache. *Headache*. 2005;45:513-9.
5. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*. 1986;51:1173-82.
6. MacCorquodale K, Meehl PE. On a distinction between hypothetical constructs and intervening variables. *Psychological Review*. 1948;55:95-107.
7. March JS, Curry JF. Predicting the Outcome of Treatment. *Journal of Abnormal Child Psychology*. 1998;26:39-51.
8. Kraemer HC, Kiernan M, Essex M, Kupfer DJ. How and why criteria defining moderators and mediators differ between the Baron & Kenny and MacArthur approaches. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*. 2008;27:S101-8.
9. Viechtbauer W. 31 Analysis of Moderator Effects in MetaAnalysis. *Best Practices in Quantitative Methods*. 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc.; 2008. p. 471-87.
10. Jenkins M. Evaluation of methodological search filters--a review. *Health information and libraries journal*. 2004;21(3):148-63.
11. White VJ, Glanville JM, Lefebvre C, Sheldon TA. A statistical approach to designing search filters to find systematic reviews: objectivity enhances accuracy. *Journal of Information Science*. 2001;27(6):357-70.
12. Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. *J Am Med Inform Assoc*. 2002;9:653-8.
13. Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ (Clinical research ed)*. 2012;344:e1553.
14. Pincus T, Miles C, Froud R, Underwood M, Carnes D, Taylor SJC. Methodological criteria for the assessment of moderators in systematic reviews of randomised controlled trials: a consensus study. *BMC medical research methodology*. 2011;11:14.
15. Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR. The RAND/UCLA appropriateness method user's manual. DTIC Document; 2001.
16. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ : British Medical Journal*. 2011;343.
17. Reuzel RP, van der Wilt GJ, ten Have HA, de Vries Robbe PF. Interactive technology assessment and wide reflective equilibrium. *Journal of Medicine and Philosophy*. 2001;26:245-61.
18. Janssen IM, Scheibler F, Gerhardus A. Importance of hemodialysis-related outcomes: comparison of ratings by a self-help group, clinicians, and health technology assessment authors with those by a large reference group of patients. *Patient Prefer Adherence*. 2016;10:2491-500.
19. Jansen SJ, Kievit J, Nooij MA, Stiggelbout AM. Stability of patients' preferences for chemotherapy: the impact of experience. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*. 2001;21:295-306.

20. Ely JW, Osheroff JA, Chambliss ML, Ebell MH, Rosenbaum ME. Answering physicians' clinical questions: obstacles and potential solutions. *J Am Med Inform Assoc.* 2005;12(2):217-24.
21. Eiring O, Landmark BF, Aas E, Salkeld G, Nylenna M, Nytroen K. What matters to patients? A systematic review of preferences for medication-associated outcomes in mental disorders. *BMJ open.* 2015;5(4):e007848.
22. Opmeer BC, de Borgie CA, Mol BW, Bossuyt PM. Assessing Preferences Regarding Healthcare Interventions that Involve Non-Health Outcomes: An Overview of Clinical Studies. *Patient.* 2010;3:1-10.
23. Janssen IM, Gerhardus A, Schroer-Gunther MA, Scheibler F. A descriptive review on methods to prioritize outcomes in a health care context. *Health Expect.* 2015;18(6):1873-93.
24. van Hoorn R, Kievit W, Booth A, Mozygemba K, Lysdahl KB, Refolo P, et al. The development of PubMed search strategies for patient preferences for treatment outcomes. *BMC medical research methodology.* 2016;16:88.
25. Janssen IM, Gerhardus A, Schröer-Günther MA, Scheibler F. A descriptive review on methods to prioritize outcomes in a health care context. *Health Expectations.* 2014.
26. Wahlster P, Brereton, L., Burns, J., Hofmann, B., Mozygemba, K., Oortwijn, W., Pfadenhauer, L., Polus, S., Rehfuess, E., Schilling, I., Van Hoorn, R., Van der Wilt, G.J., Baltussen, R., Gerhardus, A. Guidance on the integrated assessment of complex health technologies - The INTEGRATE-HTA Model. Available via: <http://www.integrate-hta.eu/downloads/> (accessed on 20st of July 2017)
27. Rohwer AB, A.; Pfadenhauer, L.; Brereton, L.; Gerhardus, A.; Mozygemba, K.; Oortwijn, W.; Tummers, M.; Van der Wilt, G.J.; Rehfuess, E. Guidance on the use of logic models in health technology assessments of complex interventions. Available via: <http://www.integrate-hta.eu/downloads/> (accessed on 20st of July 2017)
28. Brereton L, Wahlster P, Lysdahl KB, Mozygemba K, Burns J, Chilcott JB, et al. Integrated assessment of home based palliative care with and without reinforced caregiver support: 'A Demonstration of INTEGRATE-HTA methodological guidances'. 2016. Available via: <http://www.integrate-hta.eu/downloads/> (accessed on 20st of July 2017)
29. Barnato AE, Anthony DL, Skinner J, Gallagher PM, Fisher ES. Racial and ethnic differences in preferences for end-of-life treatment. *Journal of General Internal Medicine.* 2009;24:695-701.
30. Isaacs JD, Ferraccioli G. The need for personalised medicine for rheumatoid arthritis. *Annals of the rheumatic diseases.* 2011;70(1):4-7.
31. Jakka S, Rossbach M. An economic perspective on personalized medicine. *HUGO J.* 2013;7(1):1-6.
32. Bridges JF, Jones C. Patient-based health technology assessment: a vision of the future. *International Journal of Technology Assessment in Health Care.* 2007;23:30-5.

Table 1: Search filters for articles on moderators and predictors of treatment effects.

Search filter	Se (%) [95% CI]	Sp (%) [95% CI]	Ac (%)	NNR
Optimal sensitivity†				
("Epidemiologic Methods"[mesh] OR assign* OR control*[tiab] OR trial*[tiab]) AND therapy[sh]	89.1 [87.8 - 90.4]	80.2 [78.6 - 81.9]	80.6	5.59
("Epidemiologic Methods"[mesh] OR assign* OR control*[tiab]) AND (therapy[sh] OR primary*[tiab])	91.1 [89.9 - 92.3]	79.3 [77.6 - 81.0]	79.8	5.71
("Epidemiologic Methods"[mesh] OR analys* OR predict* OR trial*[tiab]) AND therapy[sh]	92.1 [90.9 - 93.2]	79.9 [78.3 - 81.6]	80.5	5.51
Optimal specificity†				
group*[tw] AND therapy*	58.4 [56.4 - 60.5]	94.9 [94.0 - 95.8]	93.2	2.81
randomi* AND treat*	61.4 [59.3 - 63.4]	94.6 [93.7 - 95.6]	93.1	2.81
group*[tw] AND treat*[tw]	65.3 [63.4 - 67.3]	94.6 [93.7 - 95.6]	93.3	2.70
Optimal accuracy†				
group*[tw] AND therapy*	58.4 [56.4 - 60.5]	94.9 [94.0 - 95.8]	93.2	2.81
(randomi* OR hazard*) AND treat*	66.3 [64.4 - 68.3]	94.3 [93.3 - 95.3]	93.0	2.78
randomi* AND treat*	61.4 [59.3 - 63.4]	94.6 [93.7 - 95.6]	93.1	2.81
Optimal NNR†				
(randomi* OR hazard*) AND treat*	66.3 [64.4 - 68.3]	94.3 [93.3 - 95.3]	93.0	2.78
(randomi* OR multivariate) AND treat*	66.3 [64.4 - 68.3]	94.1 [93.1 - 95.1]	92.8	2.85
randomi* AND (treat* OR death*)	64.4 [62.4 - 66.4]	94.4 [93.5 - 95.4]	93.1	2.78

Caption: Combinations of search terms with the best sensitivity, best specificity, and lowest NNR for detecting articles reporting on moderators or predictors of treatment outcome. Se: Sensitivity, Sp: Specificity, Ac: Accuracy, NNR: Number Needed to Read, [tiab] = title/abstract, words and numbers included in the title, collection title, abstract, and other abstract of a citation, [ti] = title, words and numbers included in the title or collection title. †Keeping sensitivity $\geq 25\%$, specificity $\geq 75\%$, and accuracy $\geq 75\%$. Mesh: Medical Subject Headings. Sh: subject heading

Table 2. Search filters for articles on patient preferences for treatment outcomes.

Search term	Se (%) [95% CI]	Sp (%) [95% CI]	Ac (%)	NNR
Best sensitivity*				
Preferen*	75.0 [73.7 - 76.3]	97.2 [96.7 - 97.7]	97.1	13.9
Relat*[tiab]	75.0 [73.7 - 76.3]	79.5 [78.2 - 80.7]	79.4	94.8
"Middle Aged"[Mesh]	66.7 [65.2 - 68.1]	77.6 [76.3 - 78.9]	77.5	116.7
Best specificity*				
Logit	33.3 [31.9 - 34.8]	99.8 [99.7 - 100.0]	99.6	2.8
"Choice Behavior" [Mesh]	33.3 [31.9 - 34.8]	99.3 [99.0 - 99.6]	99.1	8.3
"Patient Preference" [Mesh]	50.0 [48.5 - 51.5]	99.3 [99.0 - 99.5]	99.1	6.0
Best accuracy*				
Logit	33.3 [31.9 - 34.8]	99.8 [99.7 - 100.0]	99.6	2.8
"Patient Preference"[mesh]	50.0 [48.5 - 51.5]	99.3 [99.0 - 99.5]	99.1	6.0
Choice Behavior[mh]	33.3 [31.9 - 34.8]	99.3 [99.0 - 99.6]	99.1	8.3
Lowest NNR*				
Logit	33.3 [31.9 - 34.8]	99.8 [99.7 - 100.0]	99.6	2.8
"Patient Preference"[mesh]	50.0 [48.5 - 51.5]	99.3 [99.0 - 99.5]	99.1	6.0
"Choice Behavior" [mesh]	33.3 [31.9 - 34.8]	99.3 [99.0 - 99.6]	99.1	8.3

Caption: Combination of terms with the best sensitivity, best specificity, and lowest NNR. Se: Sensitivity, Sp: Specificity, Ac: Accuracy, NNR: Number Needed to Read, [tiab] = title/abstract, words and numbers included in the title, collection title, abstract, and other abstract of a citation, [ti] = title, words and numbers included in the title or collection title. *Keeping sensitivity $\geq 25\%$, specificity $\geq 75\%$, and accuracy $\geq 75\%$. Mesh: Medical Subject Headings. Sh: subject heading

Table 3: Questions in CHAMP for assessing moderators and predictors of treatment effects.

Design

1. A priori plausibility: was there sufficient empirical or theoretical support for the moderator or predictor that was examined?
2. Was the moderator or predictor specified a priori?
3. Was the moderator or predictor variable measured before the allocation or start of the intervention?
4. Was measurement of the moderator or predictor reliable and valid in the target population?

Analysis

5. In case of a moderator, was an interaction test used?
6. Was a limited number of moderators and predictors tested?
7. Was sample size adequate for the moderator or predictor analysis?

Results

8. Were results presented for all candidate moderators or predictors that were examined?
9. Did statistical tests or confidence intervals indicate that observed moderator or predictor effects were unlikely to be merely due to chance variation?
10. Was the moderator or predictor effect consistent with related moderators or predictors, or across related outcomes measured within the study?

Transferability

11. Were the setting and study population comparable to the setting and population in which the information would be used?
12. Is the moderator or predictor effect clinically important?

Body of evidence

13. Was the moderator or predictor effect consistent with related moderators or predictors, or across related outcomes measured between the studies?
14. Were the setting and study population comparable to the setting and population in which the information would be used?
15. Is the moderator or predictor effect clinically important?
16. Was the moderator or predictor effect reasonably homogenous across studies?

17. Was the moderator or predictor measured similarly across the included studies, or was an adequate conversion performed?

Note that questions 10-12 are listed twice, as they are applicable both to individual studies and sets of studies covering the same moderator or predictor (body of evidence).

Table 4: *Questions in the appraisal tool for papers on patient preferences for treatment outcomes*

Appraisal items for patient preferences for treatment outcomes

1. Does the study address relevant patient preferences for treatment outcome?
2. Is the description of population, methods, and analysis clear and complete?
3. Are the data collection methods appropriate and appropriately used?
 - a. Is the format of included questions appropriate?
 - b. Is the chosen mode of application for included questions appropriate?
4. Are any theories, assumptions or models on which the research is based adequately described?
5. Were the methods properly executed and the results reliable and valid?
6. Are the results transferable to the target population?