



Factors Limiting Subgroup Analysis in Cost-Effectiveness Analysis and a Call for Transparency

Gemma E. Shields¹ · Mark Wilberforce² · Paul Clarkson³ · Tracey Farragher⁴ · Arpana Verma^{4,5} · Linda M. Davies¹

Accepted: 18 October 2021 / Published online: 29 October 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract

The use of population averages in cost-effectiveness analysis may hide important differences across subgroups, potentially resulting in suboptimal resource allocation, reduced population health and/or increased health inequalities. We discuss the factors that limit subgroup analysis in cost-effectiveness analysis and propose more thorough and transparent reporting. There are many issues that may limit whether subgroup analysis can be robustly included in cost-effectiveness analysis, including challenges with prespecifying and justifying subgroup analysis, identifying subgroups that can be implemented (identified and targeted) in practice, resource and data requirements, and statistical and ethical concerns. These affect every stage of the design, development and reporting of cost-effectiveness analyses. It may not always be possible to include and report relevant subgroups in cost effectiveness, e.g. due to data limitations. Reasons for not conducting subgroup analysis may be heterogeneous, and the consequences of not acknowledging patient heterogeneity can be substantial. We recommend that when potentially relevant subgroups have not been included in a cost-effectiveness analysis, authors report this and discuss their rationale and the limitations of this. Greater transparency of subgroup reporting should provide a starting point to overcoming these challenges in future research.

Key Points for Decision Makers

Neglecting to consider or include subgroup analysis in cost-effectiveness analysis may mask key differences between subgroups and result in suboptimal resource allocation.

Given the quantity and range of factors limiting subgroup analysis, we encourage future researchers to be more explicit in reporting if subgroup analysis has not been presented.

Researchers and decision makers must be aware of the barriers and challenges around conducting subgroup analysis, in order to identify solutions to conduct robust subgroup analysis or to understand the potential limitations of more exploratory analysis.

✉ Gemma E. Shields
gemma.shields@manchester.ac.uk

¹ Division of Population Health, Health Services Research, and Primary Care, Faculty of Biology, Medicine and Health, Manchester Centre for Health Economics, School of Health Sciences, University of Manchester, Manchester, UK

² Social Policy Research Unit, Department of Social Policy and Social Work, University of York, York, UK

³ Social Care and Society, Division of Population Health, Health Services Research, and Primary Care, Faculty of Biology, Medicine and Health, School of Health Sciences, University of Manchester, Manchester, UK

⁴ The Epidemiology and Public Health Group (EPHG), Division of Population Health, Health Services Research, and Primary Care, Faculty of Biology, Medicine and Health, School of Health Sciences, University of Manchester, Manchester, UK

⁵ Manchester Academic Health Science Centre, University of Manchester, Manchester, UK

1 Introduction

Patient heterogeneity describes natural variation across people, which can be explained by their characteristics (including demographics, clinical characteristics and preferences) [1–3]. A subgroup is a subset of patients within a wider patient population, who are defined using one or more characteristics. Sculpher outlines the various forms of patient heterogeneity that can be used to consider subgroups, including whether factors are known at treatment, whether these are related to the treatment and/or the disease, and preferences [4]. While clinical evidence often focuses on heterogeneity in relation to treatment effect, cost-effectiveness studies need to consider wider sources of heterogeneity, e.g. related to baseline event rates [4].

Economic evaluations are only one source of information to support decision makers who must balance clinical and qualitative evaluations with policy objectives and stakeholder priorities. Cost-effectiveness analysis often uses population averages, which can hide differences between subgroups who may receive a healthcare intervention [1]. An intervention may appear cost effective across a sample but not be cost effective for one or more subgroups (e.g. if it has an unfavourable adverse effect profile, reduced efficacy, or for other reasons), and vice versa. This could result in inefficient decision making for specific subgroups, risking a suboptimal distribution of resources and unnecessary harm to patients and/or patients missing out on health benefits. Subsequently, acknowledging patient heterogeneity could increase efficiency and result in population health gains. Subgroup analyses are imperative if cost-effectiveness estimates are to reflect patient heterogeneity with the purposes of affecting decision making. However, a review identified that subgroup analyses were reported in a minority (19%) of published cost-effectiveness analysis [5]. Moreover, the review concluded that over half of the reported subgroup analyses had the potential to affect decision making (i.e. subgroups had different conclusions regarding cost effectiveness when the cost per QALY was compared with the average). However, as noted above, this would depend on the balance of priorities and objectives and broader evidence in a specific decision-making context.

Subgroups will not always be relevant (important and informative) for cost-effectiveness analysis, e.g. when regulatory bodies review trial evidence and restrict a license to an homogenous subgroup. Nevertheless, when this is not the case, there are many factors that may limit the investigation of subgroups. The paper focuses on subgroups that are meaningful for decision making, although it is recognised that subgroups may be useful for academic purposes even if they cannot be used in decision making and clinical practice (e.g. if they cannot be targeted in practice).

This paper outlines and discusses the key factors that may limit subgroup analysis in cost-effectiveness studies. Understanding these issues will be useful to researchers, reviewers, and decision makers alike. It goes on to suggest changes to the reporting of subgroup analysis, to enhance transparency and to prompt future research if relevant.

2 Deciding on a Subgroup Analysis

Focusing on subgroup analysis, the first stage in a cost-effectiveness analysis is to decide which subgroups to include; however, researchers may face many obstacles when it comes to formulating their research plan. These are discussed below.

2.1 Justification

The first step to specifying a subgroup is the choice of characteristics used to define subgroup membership [6]. Ramaekers et al. found the majority of technology appraisal guidelines require any acknowledgement of patient heterogeneity to be justified and prespecified, with biological, clinical and/or statistical reasoning [3]. Similarly, commonly used economic evaluation checklists emphasise the need to prespecify, explain and justify subgroups [7–9]. However, it has been noted that across guidelines there is a lack of clarity with respect to sources of heterogeneity that should be considered and acceptable methods and justification for subgroup analysis [3, 10, 11]. Existing publications have called for more consensus, clarity and systematic processes for exploring patient heterogeneity [3, 12].

Heterogeneity outside of economic evaluation typically focuses on treatment effect and describes how patient characteristics can be used to explain or predict different treatment effects across a population [12]. Economic evaluation has a wider range of parameters and, subsequently, the consideration of patient heterogeneity needs to extend to other parameters, including resource use, health state utility and baseline risk [1, 4]. Researchers need to think more broadly about how patient characteristics may impact the results of cost-effectiveness analysis, rather than restricting to what has been used in clinical effectiveness analysis. Given the complexity of subgroup analysis in cost-effectiveness evaluations, defining clear rules around what a subgroup analysis should be and what evidence is needed to justify it would be challenging and potentially restrictive, but this lack of clear guidance for subgroups may be off-putting to researchers. Note the lack of clear guidance is not specific to economic evaluation. Wijn et al. reviewed guidance for subgroup effects of medical treatments, which covered industry, health technology assessment agencies, academic/non-profit

organisations and regulatory bodies [13]. They found there were significant differences across the available guidance.

Building a strong justification for subgroup analysis is challenging. The thought process for defining potential subgroups is complicated, even before considering which subgroups may have expected differences in cost effectiveness. Researchers need to review a long list of potential patient characteristics and then think about interactions between characteristics and potential confounders (e.g. geographical heterogeneity). Once sources of patient heterogeneity have been considered, researchers need to consider whether and how they may influence multiple parameters of economic evaluation (e.g. baseline risk, treatment effect, resource use, utility). Not all heterogeneity will be observable when a treatment decision is made, i.e. it may be observed over time, and, subsequently, not all subgroups are informative for cost-effectiveness estimates for the purposes of decision making [4]. For example, patients may respond differently to an intervention, which cannot be predicted at the time of decision making. However, this heterogeneity, rather than informing a subgroup, can be factored into cost-effectiveness analysis by exploring stopping rules, with a recent example in chronic migraines [4, 14]. The investigation of patient heterogeneity will vary in complexity and feasibility across disease areas and intervention types. For example, there may be circumstances in which an intervention does not have a known and clear mechanism of action, which likely prevents any justification for subgroups based on treatment effect.

2.2 Prespecification

Prespecification is emphasised by guidelines; however, as noted by Sculpher, whether this can be done robustly in the early stages of work is debatable as researchers will not yet have identified all of the available evidence [4]. Fletcher et al. note that while prespecification is favoured, a more pragmatic approach is needed when it comes to cost-effectiveness analysis (e.g. when projects are waiting for phase III data, and/or conducting early-stage modelling) [15]. Furthermore, as discussed in more detail below, once subgroups have been proposed, evidence/data gaps can be problematic when conducting analysis, and therefore researchers risk prespecifying subgroups they cannot parameterise later in the research.

2.3 Ethical Concerns

Subgroup analyses can raise ethical questions, as using patient characteristics (e.g. age, sex, ethnicity) to determine access to treatment can be contentious, which may prevent them from being accepted by decision-making bodies and/or utilised in practice. Equity concerns arise if an intervention is not cost effective across all subgroups [16]. They

will vary depending on the source of heterogeneity, as well as how it affects cost effectiveness (by which parameters), the strength of the underlying justification and the potential impact on health inequalities. Whether something is ethical may depend on whether the subgroup impacts cost effectiveness via treatment effect, or by another parameter (e.g. resource use) [1, 4]. For example, Grutters et al. note that policy based on 'race' may be acceptable if it ties to biological mechanisms [1]. Furthermore, race, like other aspects of patient heterogeneity, can be defined in different ways and may be closely related to other terms that mean something different (e.g. ethnicity) [17]. Another example is sex and gender [18]. Choice of precise demographic characteristics to collect, and their definitions, is likely to affect whether the subgroup is ethical to decision makers. There are many types of subgroup that may cause concern, examples include a subgroup eligible for expensive downstream treatment that appears less cost effective because of this; subgroups that are simply based on waiting until a patient has progressed to a more severe health state; or subgroups based on protected characteristics. Espinoza et al. discussed the exclusion of age and sex from their analysis, as differentiating according to these characteristics could incur criticism related to ethics [6]. A review of national guidelines found that only the National Institute for Health and Care Excellence (NICE) lists equity constraints on subgroups (i.e. subgroups are not considered if they focus on social characteristics or location) [3, 19]. A separate review, which looked at subgroup analysis reporting in published economic evaluations, did not identify any papers discussing equity in relation to chosen subgroups [20]. The authors noted that subgroup analysis was more common in US studies, which could be reflective of reduced regulations and restrictions related to equity in decision making [20].

As noted by Petticrew et al. with respect to subgroup analysis and equity, researchers are 'damned if you do, damned if you don't' [21]. Ignoring subgroup analysis could indirectly increase inequities (e.g. if intervention reduces health in an already disadvantaged subgroup). For example, there is evidence that some public health interventions (e.g. workplace smoking bans) increase inequality, but if the average effect is favourable, these results may be hidden [22]. Methods are available to investigate the distribution of costs and outcomes in cost-effectiveness analysis and efficiency losses associated with equity constraints [23, 24].

2.4 Feasibility in Practice

Subgroups must be identifiable and targetable in practice; that is, the patient characteristics used by researchers and decision makers to define the subgroup must be available to healthcare professionals. We may have data available to us in research that can be used to identify subgroups for

academic purposes, but that might not exist in practice. Using characteristics that are routinely measured or easily observed increases the feasibility of subgroup analysis and implementation [4].

There are further complicating factors related to targeting subgroups in practice. Not all interventions can be targeted at subgroups of patients, e.g. training interventions for healthcare professionals are unlikely to be implementable for a precise subgroup only. Subgroup size, as well as affecting budget impact, may affect whether subgroup policies are adhered to in practice. For example, if a small subgroup is recommended as being unsuitable for a treatment, such guidance may not always be followed in practice. Furthermore, the number of subgroups can be important, since if there are many subgroups all with different recommendations, this can create organisational challenges, limiting feasibility. Implementing varying recommendations for subgroups across organisations may be more or less feasible depending on a range of factors, e.g. the time available to make and explain a treatment choice to a patient. External pressures on healthcare systems and periods of great change (e.g. the coronavirus disease 2019 [COVID-19] pandemic) that result in constantly changing subgroup recommendations may also be a challenge organisationally. Patients have multiple characteristics that vary simultaneously, which creates a challenge as subgroups based on multiple characteristics can be even harder to identify and justify [25]. An example of subgroups based on multiple characteristics is presented by Burn et al., who examined based on both age and sex for a cost-effectiveness analysis of total knee replacement [26]. Univariate-based subgroup analyses are more commonly reported in cost-effectiveness studies but are potentially an oversimplification as interactions between characteristics may be important [20]. Overlap between subgroups needs to be avoided (e.g. people who fit into multiple groups). When there are multiple alternative subgroup specifications being considered, Espinoza et al. propose a framework using expected net health benefit with current and perfect information to guide selection (optimal subgroup definition) [6]. Finally, Sculpher noted the importance of considering whether patients could sway measurement to meet subgroup criteria to access a treatment [4].

If there is a cost to measuring an aspect of heterogeneity, this must be factored into the cost-effectiveness analysis, and it acts as another hurdle in practice (requiring budget and resources). There may be costs associated with implementing subgroup-specific guidance even when the costs do not relate to identifying heterogeneity, which should be considered, such as in producing different guidelines for subgroups [6].

3 Implementing Subgroup Analysis

Once subgroups are agreed, researchers can start to identify the evidence needed to inform subgroup analysis and to conduct the analysis and report the results; however, there remain barriers to conducting subgroup analyses, which are outlined below. Note that some of these issues may be more or less apparent depending on the precise methods used (e.g. whether a modelling study or trial-based cost-effectiveness analysis is being conducted).

3.1 Statistical Concerns

Avoiding subgroups from ‘data dredging’ is a known issue in economic evaluation [4, 7]. Espinoza et al. discussed that health care decision making has been hesitant to adopt subgroup analysis due to statistical concerns around power and multiple testing [6]. Typically, clinical trials are powered to identify significant average treatment effects across the sample; subsequently, any subgroup analysis is likely to be underpowered. Inadequate power can result in false negatives, whereby subgroups do not appear to be important or significant due to a lack of statistical power [25]. Multiplicity becomes an issue as when multiple subgroups are compared, identifying a difference between subgroups can occur by chance (a type 1 error) [27, 28], i.e. there is a risk for false positive results. Petticrew et al. summarise historical examples of subgroups that have resulted in harmful decisions, such as limiting aspirin use for the prevention of heart disease in women [21]. Given these concerns, researchers may be cautious that evidence is not sufficient to identify subgroups, or that apparent subgroups have been identified by chance and therefore do not reflect reality. Successful replication of the subgroup results using multiple data sources has been noted as one option to increase the credibility of results; however, as noted below, there are often significant data limitations affecting analyses [15]. As factitious subgroups might arise due to these statistical issues, a robust justification and underlying rationale for subgroups is especially imperative as it can mitigate some of these concerns. Furthermore, when using estimates from meta-analysis to inform baseline risk or other parameters, researchers need to be wary of ecological fallacy [29]. This can occur when average patient characteristics are regressed against average outcomes across studies (rather than within studies), and any association found with this aggregate data may not apply to individuals within the studies.

3.2 Evidence Requirements

Stratification (dividing the potential patient group into subgroups) is needed to derive parameters for subgroup analysis. There are multiple parameters in economic evaluation that could be affected by patient heterogeneity, including resource use, health state utility, baseline risk and treatment effect [1, 4]; subsequently, conducting subgroup analysis can require a considerable amount of data.

The issues related to identifying evidence may vary depending on whether a researcher is conducting a modelling study or an analysis integrated within a trial or observational study. Subgroup analyses can be conducted irrespective of model design, although simulation models may be able to handle the complexities of heterogeneity more efficiently compared with cohort models [1]. Such models are data hungry and data availability is a concern for researchers [30]. Modelling often relies on the use of published evidence (e.g. systematic reviews and meta-analysis); however, as noted in the Cochrane handbook, insufficient details and a lack of consistency in source materials affect the feasibility of subgroup analyses in systematic reviews [31]. Selective reporting may affect the availability of evidence to inform economic evaluation. A review comparison of randomised controlled trial (RCT) protocols and articles found that 12% of publications did not report subgroup analyses that were in the relevant protocol and 26% reported subgroup analyses that were not prespecified in protocols [32]. This may reduce the credibility, and availability, of subgroup data to inform modelling.

Subgroups are more easily investigated when patient-level data are available. However, economic evaluations integrated within RCTs are often limited by sample size and are seldom powered for economic outcomes (linked to the statistical concerns outlined above). This was the case in an economic evaluation within a trial reported by Wijnen et al., which failed to recruit the target number of participants and loss to follow-up added to the sample size and power problems [33]. The authors did not identify any subgroup effects; however, this may be indicative of an insufficient sample size rather than a lack of significant subgroups. Hoch et al. investigated the cost effectiveness of assertive community treatment and found differences in cost effectiveness between subgroups based on race, and differences in uncertainty estimates between subgroups that were attributed to subgroup sample sizes [34]. Assessing the level of uncertainty is key for decision making and limited sample sizes will contribute to parameter uncertainty [35].

Strict trial inclusion/exclusion criteria can also limit the usefulness of trials with regard to subgroup analysis, as they may restrict to a more homogenous population, e.g. Marshall and Hux discussed that RCTs for coxibs typically exclude patients with cardiovascular disease, despite evidence to

suggest this pertains to 40% of patients in practice [36]. Post-launch, real-world data may offer some advantages for subgroup analysis in cost-effectiveness studies, as wider populations are considered and subgroup analysis can reflect groups targeted in practice. A recent example in asthma demonstrates the use of real-world data to inform subgroups beyond trial data (with a focus on older populations) [37]. Research used to inform economic evaluation (e.g. trials or observational studies) can only collect a limited number of patient characteristics. Sculpher and Gafni note this is an issue in identifying preference heterogeneity, as the range of sociodemographic data collected can be limited [38]. Additionally, health economists will not always be involved in the design of data collection, which may limit the identification of patient heterogeneity that is specifically relevant to economic evaluation. Further complicating factors exist. For example, self-reported data may be subject to response bias, which may differ according to participant characteristics [39]. Note, these issues will also affect modelling studies as they reduce the evidence base available to populate a model.

Conducting subgroup analyses in economic evaluations increases data requirements and, subsequently, may not always be feasible depending on the evidence base. While there are some ideas in the literature that could reduce issues related to insufficient data (e.g. open data policies), realistically there will always be some limit to this due to research constraints (including funding and resources) [12, 40].

3.3 Analysis and Reporting

Grutters et al. systematically reviewed methods to acknowledge patient heterogeneity in cost-effectiveness analysis [1]. Various methods are available for this, including regression techniques, adaptations to modelling, and value of information methods that can be used when data limitations are an issue [1, 2]. Choosing appropriate methods and reflecting wider guidance on methods (e.g. from health technology assessment bodies) is a further challenge for researchers. Furthermore, reporting subgroup analyses can be time intensive, especially if full results (including cost-effectiveness acceptability curves) are presented and multiple subgroups are considered [1, 2].

Arguably, selective reporting of subgroups in published cost-effectiveness analysis often cannot be identified, as protocols for economic evaluation are not routinely published. The process of defining subgroups may be more transparent in technology appraisal if bodies place an emphasis on robust subgroup identification and reporting. NICE are currently conducting a review and consultation of their methods of health technology appraisal and in the case for change describe that committees must be able to exclude subgroups for whom a technology is not cost effective even when it appears cost effective across the whole wider population

[41]. While this is imperative for efficiency and for revealing true population health gains, in circumstances in which an intervention appears cost effective across a whole population, there may currently be little incentive to investigate subgroups in which treatment is not cost effective. For instance, manufacturers funding an economic evaluation may have an interest in positive outcomes that will maximise (prioritise) profit, and researchers without conflicts of interest may still be keener to publish favourable results [42]. Fletcher et al. present an example of a health technology assessment case study in Alzheimer's disease to demonstrate the importance of subgroup analyses in cost-effectiveness analyses [15]. Although it cannot be evidenced, media reporting of health technology assessment recommendations adds extra complexity, as recommendations restricted to a particular subgroup of the patient population may be unpopular as it may be judged to be inequitable or unfair.

4 Discussion

Heterogeneity in economic evaluation is complex, with multiple patient characteristics and parameters of economic evaluation to consider. It has previously been identified that only a minority of cost-effectiveness analyses report subgroup analysis and this paper examines some of the key obstacles facing researchers [20]. This includes issues with prespecifying and justifying subgroup analysis, identifying subgroups that can be implemented in practice, resource and data requirements, statistical concerns, and ethical concerns. Figure 1 simply summarises the factors discussed here. The reported challenges related to subgroup analysis are likely to vary across populations and specific decision-making context.

Even if subgroup analysis cannot be conducted, e.g. due to evidence requirements, the consequences of making incorrect decisions because of this should be considered. It is often stated that the objective of economic evaluation

is to maximise population health for a given budget, and, subsequently, neglecting subgroup analyses may prevent this objective from being achieved.

Although existing checklists emphasise the need to pre-specify, explain, and justify subgroups, they do not address subgroups that were omitted [7–9]. To allow the evidence base to be improved over time, it would be useful for other researchers to understand why subgroup analysis has not been conducted and whether if, in future, this would be useful and if any obstacles could be reduced (e.g. through data collection). This would also allow readers to understand any limitations related to missing subgroup analysis. For these reasons, we call for increased transparency in subgroup reporting, with attention given to what was not reported and why, with the aim of improving future research and more thoroughly understanding the limitations of current studies. We would encourage researchers to consider and respond to the questions included in Table 1.

Alongside increased transparency, we recommend that researchers consider potential subgroups early in the process of economic evaluation and systematically identify subgroups, using techniques such as logic models and causal inference, to define subgroups and complex interactions between sources of heterogeneity and outcomes. Increased stakeholder engagement throughout the process of economic evaluation (from conceptualisation to final reporting) may also support the identification and justification of subgroup analysis. An example of stakeholder engagement in guiding economic model development and subgroups (including identifying subgroups and relevant data sources) has been reported by Xie et al. [43]. We would also encourage researchers to consider the full range of methods available to acknowledge patient heterogeneity in cost-effectiveness analysis (not restricted to subgroup analysis), as well as any guidelines that are applicable to their setting of interest [1].

We recognise that subgroup analysis is not always helpful or feasible but encourage researchers to be transparent about their thought processes. Without knowing whether

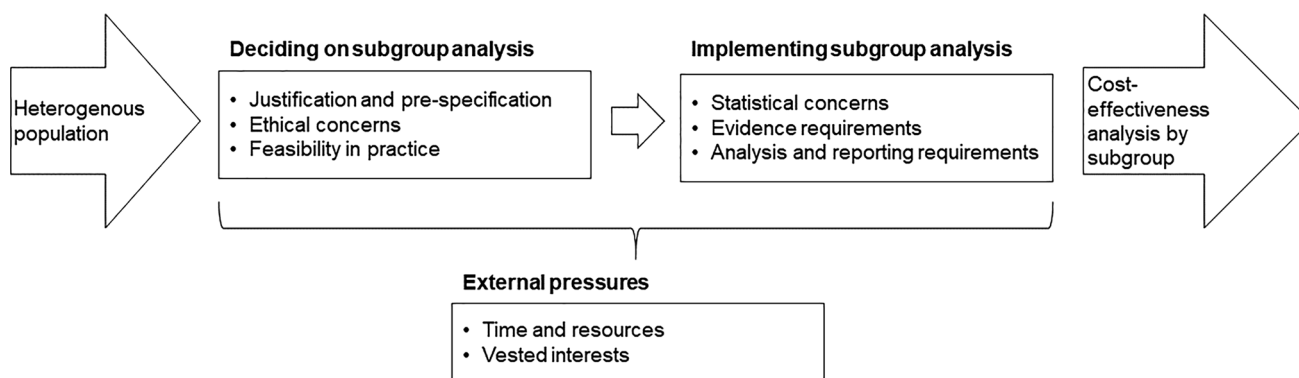


Fig. 1 Overview of subgroup considerations

Table 1 Questions to consider for transparent subgroup reporting

Number	Question
1	Were any subgroups considered but excluded from the cost-effectiveness analysis?
If answered yes to the above, respond to the questions below for each potential subgroup	
1.1	How was the subgroup defined and justified (i.e. biological, clinical and/or other rationale)?
1.2	Which parameters of the economic evaluation would the subgroup change (e.g. baseline rates, resource use, utility, treatment effect)?
1.3	What was the potential direction of the subgroup analysis (i.e. evaluating the existing evidence, would intervention potentially be cost effective or not cost effective in this group)?
1.4	Could future research improve the feasibility of this subgroup analysis in this area and if so, how?

subgroups were considered and, if so, why they were ruled out, other researchers in the area will be unable to learn from this and decision makers may fail to recognise a key limitation of the evidence base. When subgroup analyses are reliant on building a stronger/more comprehensive evidence base, we hope that increased transparency in reporting will help to achieve this.

Declarations

Funding The authors received no financial support for the research, authorship, or publication of this article.

Conflicts of interest/Competing interests Gemma E. Shields, Mark Wilberforce, Paul Clarkson, Tracey Farragher, Arpana Verma and Linda M. Davies have completed the International Committee of Medical Journal Editors (ICMJE) uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare no support from any organisation for the submitted work (with the exclusion of the stated funder), no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years, and no other relationships or activities that could appear to have influenced the submitted work.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and material Not applicable.

Code availability Not applicable.

Author contributions Conceptualisation: GES, MW, PC, TF, AV and LMD. Writing—original draft preparation: GES. Writing—review and editing: GES, MW, PC, TF, AV and LMD.

References

- Grutters JPC, Sculpher M, Briggs AH, et al. Acknowledging patient heterogeneity in economic evaluation. *Pharmacoeconomics*. 2013;31:111–23. <https://doi.org/10.1007/s40273-012-0015-4>.
- Briggs A, Sculpher M, Claxton K. *Decision modelling for health economic evaluation*. Oxford: Oxford University Press; 2006.
- Ramaekers BLT, Joore MA, Grutters JPC. How should we deal with patient heterogeneity in economic evaluation: a systematic review of national pharmacoeconomic guidelines. *Value Health*. 2013;16:855–62. <https://doi.org/10.1016/J.JVAL.2013.02.013>.
- Sculpher M. Subgroups and heterogeneity in cost-effectiveness analysis. *Pharmacoeconomics*. 2008;26:799–806. <https://doi.org/10.2165/00019053-200826090-00009>.
- Lavelle TA, Kent DM, Lundquist CM, et al. Patient variability seldom assessed in cost-effectiveness studies. *Med Decis Mak*. 2018;38:487–94. <https://doi.org/10.1177/0272989X17746989>.
- Espinoza MA, Manca A, Claxton K, et al. The value of heterogeneity for cost-effectiveness subgroup analysis: conceptual framework and application. *Med Decis Mak*. 2014;34:951–64. <https://doi.org/10.1177/0272989X14538705>.
- Husereau D, Drummond M, Petrou S, et al. Consolidated health economic evaluation reporting standards (CHEERS) statement. *BMJ*. 2013;346: f1049. <https://doi.org/10.1136/bmj.f1049>.
- Philips Z, Bojke L, Sculpher M, et al. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics*. 2006;24:355–71. <https://doi.org/10.2165/00019053-200624040-00006>.
- Chiou CF, Hay JW, Wallace JF, et al. Development and validation of a grading system for the quality of cost-effectiveness studies. *Med Care*. 2003;41:32–44. <https://doi.org/10.1097/00005650-200301000-00007>.
- Faria R. Problem structuring in economic evaluation. In: *Oxford research encyclopedia of economics and finance*. Oxford: Oxford University Press; 2021. doi:<https://doi.org/10.1093/acrefore/9780190625979.013.108>
- Zisis K, Naoum P, Athanasakis K. Qualitative comparative analysis of health economic evaluation guidelines for health technology assessment in European countries. *Int J Technol Assess Health Care*. 2021;37: e2. <https://doi.org/10.1017/S0266462320002081>.
- Espinoza MA, Sculpher MJ, Manca A, et al. Analysing Heterogeneity to Support Decision Making. In: *Encyclopedia of Health Economics*. Elsevier; 2014. pp. 71–6. doi:<https://doi.org/10.1016/B978-0-12-375678-7.01420-6>
- Wijn SRW, Rovers MM, Le LH, et al. Guidance from key organisations on exploring, confirming and interpreting subgroup effects of medical treatments: A scoping review. *BMJ Open*. 2019;9: e028751. <https://doi.org/10.1136/bmjopen-2018-028751>.
- Hollier-Hann G, Curry A, Onishchenko K, et al. Updated cost-effectiveness analysis of onabotulinumtoxinA for the prevention of headache in adults with chronic migraine who have previously received three or more preventive treatments in the UK. *J Med Econ*. 2020;23:113–23. <https://doi.org/10.1080/13696998.2019.1675417>.

15. Fletcher C, Chuang-Stein C, Paget M-A, et al. Subgroup analyses in cost-effectiveness analyses to support health technology assessments. *Pharm Stat.* 2014;13:265–74. <https://doi.org/10.1002/pst.1626>.
16. Culyer AJ, Bombard Y. An equity framework for health technology assessments. *Med Decis Mak.* 2012;32:428–41. <https://doi.org/10.1177/0272989X11426484>.
17. Lee C. “Race” and “ethnicity” in biomedical research: How do scientists construct and explain differences in health? *Soc Sci Med.* 2009;68:1183–90. <https://doi.org/10.1016/J.SOCSCIMED.2008.12.036>.
18. Clayton JA, Tannenbaum C. Reporting Sex, Gender, or Both in Clinical Research? *JAMA.* 2016;316:1863–4. <https://doi.org/10.1001/JAMA.2016.16405>.
19. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. London: National Institute for Health and Care Excellence; 2013.
20. Lavelle TA, Kent DM, Lundquist CM, et al. Patient variability seldom assessed in cost-effectiveness studies. *Med Decis Making.* 2018;38:487–94. <https://doi.org/10.1177/0272989X17746989>.
21. Petticrew M, Tugwell P, Kristjansson E, et al. Damned if you do, damned if you don't: subgroup analysis and equity. *J Epidemiol Community Health.* 2012;66:95–8. <https://doi.org/10.1136/jech.2010.121095>.
22. Lorenc T, Petticrew M, Welch V, et al. What types of interventions generate inequalities? Evidence from systematic reviews. *J Epidemiol Community Health.* 2013;67:190–3. <https://doi.org/10.1136/jech-2012-201257>.
23. Angus C, Cookson R, Griffin S, et al. Distributional cost-effectiveness analysis: quantifying health equity impacts and trade-offs. Oxford: Oxford University Press; 2020.
24. Coyle D, Buxton MJ, O'Brien BJ. Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. *Health Econ.* 2003;12:421–7. <https://doi.org/10.1002/hec.788>.
25. Burke JF, Sussman JB, Kent DM, et al. Three simple rules to ensure reasonably credible subgroup analyses. *BMJ.* 2015;351:h5651. <https://doi.org/10.1136/bmj.h5651>.
26. Burn E, Liddle AD, Hamilton TW, et al. Cost-effectiveness of unicompartmental compared with total knee replacement: a population-based study using data from the National Joint Registry for England and Wales. *BMJ Open.* 2018;8:e020977. <https://doi.org/10.1136/bmjopen-2017-020977>.
27. Ranganathan P, Pramesh C, Buyse M. Common pitfalls in statistical analysis: the perils of multiple testing. *Perspect Clin Res.* 2016;7:106. <https://doi.org/10.4103/2229-3485.179436>.
28. Nixon RM, Thompson SG. Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. *Health Econ.* 2005;14:1217–29. <https://doi.org/10.1002/hec.1008>.
29. Da Costa BR, Sutton AJ. A comparison of the statistical performance of different meta-analysis models for the synthesis of subgroup effects from randomized clinical trials. *BMC Med Res Methodol.* 2019;19:198. <https://doi.org/10.1186/s12874-019-0831-8>.
30. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ.* 2006;15:1295–310. <https://doi.org/10.1002/hec.1148>.
31. Higgins JPT, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011]. The Cochrane Collaboration; 2011.
32. Kasenda B, Schandelmaier S, Sun X, et al. Subgroup analyses in randomised controlled trials: Cohort study on trial protocols and journal publications. *BMJ.* 2014;349:g4539. <https://doi.org/10.1136/bmj.g4539>.
33. Wijnen BFMM, Pos K, Velthorst E, et al. Economic evaluation of brief cognitive behavioural therapy for social activation in recent-onset psychosis. *PLoS ONE.* 2018;13:e0206236. <https://doi.org/10.1371/journal.pone.0206236>.
34. Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: A framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ.* 2002;11:415–30. <https://doi.org/10.1002/hec.678>.
35. Briggs AH, Weinstein MC, Fenwick EAL, et al. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Med Decis Mak.* 2012;3(2):722–32. <https://doi.org/10.1177/0272989X12458348>.
36. Marshall DA, Hux M. Design and analysis issues for economic analysis alongside clinical trials. *Med Care.* 2009;47(7 Suppl 1):S14–20. <https://doi.org/10.1097/mlr.0b013e3181a31971>.
37. Hong SH, Cho JY, Kim TB, et al. Cost-effectiveness of tiotropium in elderly patients with severe asthma using real-world data. *J Allergy Clin Immunol Pract.* 2021;9:1939–1947.e7. <https://doi.org/10.1016/J.JAIP.2020.11.052>.
38. Sculpher M, Gafni A. Recognizing diversity in public preferences: the use of preference sub-groups in cost-effectiveness analysis. *Health Econ.* 2001;10:317–24. <https://doi.org/10.1002/hec.592>.
39. Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. *Int J Behav Health Res.* 2011;2:320. <https://doi.org/10.1504/ijbhr.2011.043414>.
40. Ramaekers B, Grutters J, Joore M. Acknowledging patient heterogeneity in health technology assessment. *OA Epidemiol.* 2014;2(1):3.
41. National Institute for Health and Care Excellence. Reviewing our methods for health technology evaluation: consultation. 2020. <https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/chte-methods-consultation>. Accessed 2 Feb 2021.
42. Bell CM, Urbach DR, Ray JG, et al. Bias in published cost effectiveness studies: systematic review. *BMJ.* 2006;332:699–703.
43. Xie RZ, deFur ME, Linthicum MT, et al. Putting stakeholder engagement at the center of health economic modeling for health technology assessment in the United States. *Pharmacoeconomics.* 2021;39:631–8. <https://doi.org/10.1007/S40273-021-01036-3>.