PharmacoEconomics (2025) 43:601–616
https://doi.org/10.1007/s40273-025-01478-z

REVIEW ARTICLE

A Review of Heterogeneity in Comparative Economic Analysis, with Specific Considerations for the Decentralized US Setting and Patient-Centered Care

Michael S. Willis[1] · Andreas Nilsson[1] · Cheryl A. Neslusan[2]

Accepted: 18 February 2025 / Published online: 8 March 2025
© The Author(s) 2025

Abstract

Patient-centered care emphasizes individual preferences, but insurer coverage decisions—based on population-level evidence—may restrict treatment options for patients who differ from the average. This highlights the importance of considering heterogeneity, which refers to differences in health and cost outcomes that are systematically linked to variations in factors like patient characteristics, insurer policies, and provider practices. Failing to account for heterogeneity in economic evaluations can lead to suboptimal decisions, inferior outcomes, and inefficiency. This study aimed to assess the tools and methods for addressing heterogeneity in economic evaluations, examine the extent to which, and how, heterogeneity has been addressed in US cost-utility studies, and provide insights and recommendations to promote more fuller consideration of heterogeneity in US economic evaluations. We reviewed and adapted a seminal taxonomy of heterogeneity to the US setting, highlighting key drivers like patient preferences and insurance design. Methods for addressing heterogeneity in economic evaluations were also reviewed and summarized. We used data from the Tufts Medical Center Cost-Effectiveness Analysis Registry to assess empirical practices in US cost-utility applications, specifically the frequency, types, and impact of a subgroup analysis, and whether rationales for including or excluding subgroups were provided. The revised taxonomy highlights key drivers of heterogeneity in the diverse and decentralized US healthcare ecosystem, such as the diversity of patient preferences and in non-patient factors like access to healthcare providers and insurance coverage. Methods to explore, confirm, and incorporate heterogeneity into a comparative economic analysis exist, but are often challenged by data availability. In addition to the trade-off between potential efficiency gains and increasing uncertainty in comparative value estimates, ethical implications of stratified decisions were highlighted in the literature. We found that a subgroup analysis was rare, and primarily performed for clinical factors like age and disease severity. Only 2 of the 85 studies published between 2015 and 2022 with subgroup-level results were found to consider non-patient factors, and none considered preferences. One-third of studies reported incremental cost-effectiveness ratios differing by more than 50% from the unstratified estimate. No studies provided a rationale for omitting a subgroup analysis, and only two motivated inclusion of a subgroup analysis, limiting our ability to assess the appropriateness of these decisions. Despite well-documented methods to address heterogeneity, its application is limited in US cost-utility studies, especially regarding patient preferences and non-patient factors. As these factors often drive real-world health outcomes and costs in the USA, proper consideration of, and reporting on, heterogeneity is essential to avoid erroneous market access decisions, suboptimal patient outcomes, and economic inefficiency. Future efforts, including work by an upcoming Professional Society of Pharmacoeconomics and Outcomes Research Task Force, should continue to refine taxonomies and emphasize the importance of addressing heterogeneity.

1 Introduction

Heterogeneity, in the context of a comparative economic analysis, has been defined as the "variability in health and cost outcomes between individuals receiving the same treatment that can be explained by variability in the patient population" [1]. For example, health and cost outcomes vary systematically based on patient-specific factors such as gene expression, treatment response, disease prognosis, quality of life, preferences, and access to care [1, 2]. Features of healthcare systems can also impact health and cost outcomes systematically, independent of patients; for instance, surgical success varies with surgeon experience [3] and healthcare

Extended author information available on the last page of the article

Key Points for Decision Makers

Many patients may not receive optimal treatment because payers base coverage decisions on population averages for treatment benefits, harms, and costs. Failure to account for heterogeneity in these outcomes can lead to denial of appropriate care to patients whose needs diverge from population norms.

Most US comparative economic analyses have not accounted for heterogeneity with subgroup analyses. Among studies in our sample that did, one third reported value estimates differing by over 50% from population averages, suggesting that neglecting heterogeneity may lead to inaccurate assessments of treatment value.

Researchers should address heterogeneity comprehensively in comparative economic analyses, tailoring studies to specific treatments, populations, and healthcare ecosystems. If addressing heterogeneity is not appropriate for ethical reasons or feasible, the rationale for omitting a subgroup analysis should be clearly justified and potential implications explained.

costs vary widely across countries [4, 5], regions (e.g., urban vs rural areas) [5, 6], and care settings (e.g., primary, secondary, and tertiary facilities) [7].

While optimizing care by tailoring treatments to each patient's unique characteristics, needs, and preferences by accounting explicitly for these diverse sources of heterogeneity is a goal of patients and physicians, the treatment options available to them can be restricted by insurance coverage decisions that are based on population-level evidence (e.g., average treatment efficacy observed in trials and cost effectiveness evaluated for representative patients) [8]. These "one-size-fits-all" coverage decisions can result in suboptimal care for patients who fall outside population norms, potentially denying them access to treatments that—while perhaps not cost-effective for average patients—may be safe, effective, and of value for them individually [8, 9]. Importantly, the Second Panel on Cost-Effectiveness in Health and Medicine concluded that "when there is heterogeneity in estimates of cost-effectiveness between different subpopulations, the analyst should present these differences as subgroup analyses" [10]. Medical advances, particularly in precision diagnostics and targeted therapies, serve to heighten the importance of addressing heterogeneity [11].

Stratified coverage decisions, which use subgroup-level results to reflect heterogeneity, can help align insurance terms with the needs of specific patient subgroups and promote patient-centered outcomes [1, 2, 12]. For example, consider the decision to fund a low-dose computed tomography screening program for lung cancer. One study found an incremental cost-effectiveness ratio (ICER) of $81,000 for *ever* smokers, but there was substantial variation across subgroups of *current* smokers (50% lower at $43,000) and *former* smokers (nearly eight times higher at $615,000) [13]. An unstratified decision to fund the program for the full population of *ever* smokers based on these estimates could lead to suboptimal outcomes for one subgroup. For instance, if willingness-to-pay falls between $43,000 and $81,000, these results suggest the program should not be funded, and *current* smokers would be denied access to valuable screening. Conversely, if willingness to pay exceeds $81,000 but is less than $615,000, these results suggest the program should be funded, even though resources expended screening *former* smokers would be more efficiently spent elsewhere. Stratified and unstratified decisions only align when willingness to pay is less than $43,000 (when the program is not cost effective for either subgroup) or greater than $615,000 (when it is cost effective for both subgroups).

Despite clear implications for economic efficiency, there is limited evidence on how consistently heterogeneity is addressed in economic evaluations. The only study to examine this question found that relatively few studies in a random sample reported subgroup results—fewer than 20% of 200 cost-utility studies published in 2014 [14]. Note that the authors did not investigate why the remaining 80% of studies omitted a subgroup analysis. While there may be justifiable reasons for omission—such as careful analysis revealing limited impact or ethical concerns, like stratified decisions disadvantaging vulnerable populations [1, 2, 10, 15]—there may also be less justifiable reasons, such as unfamiliarity with best-practice recommendations on addressing heterogeneity [10, 15–17] or limited knowledge of the methods available for incorporating it into an economic analysis. Failing to consider heterogeneity in decision making, moreover, can disadvantage patients who differ from norms, which may itself be unethical. Given that a study may involve multiple relevant stratification factors, it is even possible that some of the studies reporting subgroup analyses overlooked other relevant sources of heterogeneity. Analysis of author rationales for including or excluding a subgroup analysis, the stating of which is a recommended part of study reporting [12], could provide valuable insights.

To help researchers identify and understand the full scope of variability in economic evaluations, Sculpher classified factors that may contribute to heterogeneity in health and cost outcomes into the following categories: 1. factors potentially known at treatment selection that are related to the treatment; 2. factors potentially known at treatment selection related to the disease but not to the treatment; 3.

factors potentially known at treatment selection unrelated to the disease; 4. factors potentially known at treatment selection unrelated to the patient; 5. preferences; and 6. factors revealed over time. He also provided examples to raise awareness of the potentially numerous factors that vary systematically and impact health and cost outcomes [2]. As the paper was originally prepared to support an update of the National Institute for Health and Care Excellence Guide to the Methods of Technology Appraisal for the single-payer UK setting, understandably none of the provided examples emphasize sources of heterogeneity especially pertinent in the decentralized multi-centric US healthcare system, like variability in insurance. Not only are there multiple insurers with heterogeneous case mixes, but there is also considerable variation in coverage and payment terms across the different plans offered by each insurer (e.g., differences in which treatments are excluded, deductible limits, and coinsurance rates). Heterogeneity in patient preferences regarding health states and treatment attributes may be relatively more important in the USA where cost sharing imposes significant financial considerations, and individuals often choose among insurer plans with varying terms. The extent of this financial outlay is a function of their choice of insurance plan, and as such, patient preferences regarding risk should also be considered.

The objective of this study is to enhance understanding and promote incorporation of heterogeneity in economic evaluations, with a particular focus on addressing the specific needs of the US healthcare ecosystem. To achieve this, we: (1) tailored Sculpher's informal taxonomy of heterogeneity sources to the US setting, providing examples; (2) synthesized empirical methods and analytical frameworks available for addressing heterogeneity; and (3) assessed empirical practices in US applications by examining the classes of heterogeneity considered and whether authors provided clear rationales for the inclusion or exclusion of such analyses.

2 Key Considerations for a Stratified Comparative Economic Analysis

Incorporating variability in economic evaluations requires a structured approach to ensure that differences in patient characteristics and healthcare environments—specifically those that lead to variations in health and cost outcomes for patients receiving the same treatment—are systematically accounted for. The process begins by identifying sources of heterogeneity that are relevant based on potential clinical, economic, and contextual significance, and then assessing these candidates for inclusion using both quantitative and qualitative methods. For analyses relying on modeling, rather than patient-level trial or real-world data (RWD) alone, this step also includes generating the parameters necessary to capture the heterogeneity, such as subgroup-specific treatment effects or cost estimates. The economic analysis is then performed, integrating these factors. Analytical frameworks can also be applied to estimate the expected net benefit of stratified decisions, or, conversely, the expected opportunity cost of not implementing them.

2.1 An Informal Taxonomy of Potentially Relevant Sources of Heterogeneity

The six classes of patient and non-patient heterogeneity described in Sculpher's seminal informal taxonomy capture a broad spectrum of factors that should be considered when planning economic evaluations [2]. In Table 1, we present a revised informal taxonomy, building on Sculpher's framework and incorporating several revisions suggested by Kohli-Lynch and Briggs [1].[1] The first five classes are, at least potentially, knowable at the time of treatment selection, while the sixth is not. The following paragraphs explain these classes.

The first class, "intervention-related," includes factors that vary across patients and relate to the treatment [2]. Relative treatment effects, whose variation is often explored in a subgroup analysis of clinical trial data, fall in this class. For example, statin therapy has been shown to produce greater relative risk reductions for cardiovascular events in patients with higher baseline low-density lipoprotein (LDL) cholesterol levels than in those with lower levels. In one study, patients with baseline LDL cholesterol exceeding 160 mg/dL were observed to experience twice the proportional risk reduction in cardiovascular death compared with those with LDL levels below 100 mg/dL [19]. In the US context, addressing treatment effect heterogeneity due to systematic differences in social determinants of health, like health-related behaviors and socioeconomic and environmental factors, is likely to be important, given their strong association with disparities in access to care health outcomes, and costs [20, 21]. This class also

[1] These revisions relate primarily to avoiding overlap in some class definitions. First, Kohli-Lynch and Briggs recognized that factors can logically be both "related to the intervention" (Class 1) and "unrelated to the disease" (Class 2). For example, nausea, a common side effect of glucagon-like peptide-1 receptor agonists used in diabetes mellitus treatment, is unrelated to the disease but can limit treatment adherence and, consequently, efficacy [18]. To address this overlap, they redefined Class 3 as "factors unrelated to both the disease and the intervention", assigning factors unrelated to the disease but related to the intervention exclusively to Class 1. Second, in addition to patient preferences, economic evaluations also consider non-patient preferences, such as those of healthcare providers, employers, or society (e.g., in the form of community-based utility values),which are both "unrelated to the patient" and "preferences". To resolve this

Table 1 Classes relevant for a comparative economic analysis

Heterogeneity class[a]	Examples
(1) Intervention-related factors	Statin therapy produces greater *relative risk reductions* for CVD in subgroups of patients with higher baseline LDL-C values [25] Weight-based dosing (e.g., insulin) introduces systematic cost differences [1]
(2) Disease- but not intervention-related factors	Underlying disease features like disease severity can systematically drive differences in *absolute risk* Observable patient characteristics like age and sex may systematically drive absolute risks [27] and cost differences [28]
(3) Factors unrelated to both intervention and disease	The elderly generally face greater risks risk of developing chronic diseases (e.g., CVD and cancer) and may have less capacity to benefit from intervention [1] Patients living in rural areas that experience an acute myocardial infarction may have a higher risk that events are fatal [29] Age, chronic disability, and poverty can impact costs through the type of coverage (e.g., US Medicare eligibility begins generally at age 65 years, but includes exceptions for long-term disabled individuals, and US Medicaid eligibility is based generally on income and asset thresholds, varying by state and household size)
(4) Patient preferences	Preferences for health can vary systematically with age [88] Patients in the USA value and choose among competing health insurers (including no insurance) [38]
(5) Non-patient factors	Clinician training and experience can impact surgical success rates [3] Valuation of treatments outcomes can vary across institutional payers with diverse goals and responsibilities
(6) Factors revealed *ex post*	Sufficient weight loss at a fixed timepoint may be used to determine stopping rules for weight loss products [45, 46] Improved CV prevention when the dose or choice of statin is modified to avoid common side effects

CV cardiovascular, *CVD* cardiovascular disease, *LDL-C* low-density lipoprotein cholesterol

[a]Based on work by Sculpher [2] and Kohli-Lynch and Briggs [1], but modified to better reflect the US setting

includes treatment uptake rates. Uptake will likely vary when treatment effects are heterogenous, with greater use in subgroups of patients expected to benefit disproportionally [22]. Costs that relate systematically to interventions may also be contained in this class. For instance, insulin and some chemotherapies are dosed according to patient weight [23, 24].

The second class, "disease- but not intervention-related" includes factors that vary across patients and are unrelated to the treatment. This class includes heterogeneity in baseline absolute risks related to patient characteristics. Note that even when relative risk reductions are identical in two subgroups, an intervention will prevent more events in a subgroup with a higher baseline risk. Continuing with the statin example from Class 1, higher baseline LDL levels have also been shown to be associated with a greater absolute risk [19, 25]. The finding that statin therapy prevents more events in patients with higher LDL, thus, is due to both a higher baseline risk and larger relative risk reductions, and appropriately addressing heterogeneity requires considering both. A recent example that included both classes found that the ICER was eight times higher for the subgroup with LDL levels below 3.4 mmol/L compared with the subgroup with LDL levels above this threshold [26]. Another example involving statin therapy also considered both Class 1 factors (age-specific statin initiation rates) and Class 2 factors (age- and sex-specific risks for myocardial infarction) [27]. This study also considered age- and sex-specific utility weights, which are a member of Class 4 (described below). Social determinant of health factors may also be pertinent members of this class, as they may be important determinants of disease severity. Class 2 also includes factors that relate systematically to costs; for instance, annual costs in individuals with coronary artery disease have been found to vary with age and co-morbidities like heart failure and diabetes mellitus [28].

Class 3, "unrelated to both the disease and the intervention," includes factors that influence a patient's capacity to benefit from an intervention, independent of both relative and absolute treatment effects. For example, the elderly and

Footnote 1 (continued)

overlap, non-patient preferences were reassigned to Class 4 and Class 5 was renamed "patient preferences. Third, heterogeneity in quality of life is an example of both "disease- but not intervention-related" factors and preferences. [1, 2] To reduce confusion, we assign this under "patient preferences" and "unrelated to patients", as appropriate. To consolidate the four classes where patient-related factors could be categorized, "patient preferences" were placed in Class 4, while factors "unrelated to the patient" were assigned to Class 5.

terminally ill have shorter life expectancies than others, limiting the time over which treatment-related health gains can accrue. Note that stratified decisions can raise ethical concerns, such as the potential to discriminate against the elderly or infirm (discussed further in Sect. 3) [2]. Age also impacts health outcomes indirectly through its role in determining Medicare eligibility, which affects access to care and treatment options. Socioeconomic status, proximity to care facilities [29], and individual insurance coverage terms can also impact both health outcomes and costs. Systematic differences in future costs unrelated to the disease and the intervention, such as the lifetime costs for a child surviving an otherwise fatal condition, are included in this class, and the Second Panel on Cost-Effectiveness in Health and Medicine recommends their consideration in a cost-effectiveness analysis when survival is likely to differ across interventions [10].

Class 4, "patient preferences," includes factors that relate to the value patients assign to their health and healthcare (e.g., treatment options, associated side effects, and health outcomes) [1, 2]. For example, some patients with late-stage cancer may refuse chemotherapy and opt for palliative care, suggesting a higher relative value of quality of life over quantity, while others may choose treatment. Subgroup analyses informed by these preferences can ensure that patients with varying priorities receive treatments aligned with their individual needs. In traditional cost-effectiveness analyses, the focus is on health-related quality of life and preferences for different health states are measured using survey instruments [30, 31]. Variation in how patients value health naturally results in differences in cost effectiveness. The Generalized Risk-Adjusted Cost-Effectiveness (GRACE) approach, a recent major advance over a traditional analysis, explicitly accounts for the diminishing marginal utility of health, and in so doing, variation in the preferences for health improvements [32]. Specifically, unlike the traditional approach, the value of a given health improvement is greater for patients in poorer than in better health with GRACE. This approach additionally allows for variability in preferences for risk, so that treatments offering more certainty in beneficial outcomes are valued more by those patients who are more risk averse (and vice versa for risk takers) [33]. As such, the GRACE framework can be used to consider other uncertainty-related concepts—and Class 4 factors—introduced by ISPOR's Special Task Force on Defining Elements of Value, like value of hope (the preference for treatments with a positively skewed distribution of outcomes); the value of knowing (e.g., the value of accurate predictions as to who will respond to treatment); and insurance value (value to healthy individuals of being protected from both the physical and financial burden of a disease) [34–36]. In the USA, preferences regarding the financial outlay required to receive a treatment may be particularly pertinent. Accounting for the diversity in the willingness to pay for treatments can inform the design of insurance terms that encourage patients receiving care that is of high value to them [37, 38].

The defining feature of Class 5, "non-patient factors," is systematic variability unrelated to the patient. In the USA, with a diversity of stakeholders and stakeholder goals, there is an especially high degree of variation in the key non-patient drivers of health and cost outcomes [39]. Studies have shown that outcomes vary with healthcare system features like hospital characteristics and insurance type. For example, the mortality risk for patients admitted for acute myocardial infarction or heart failure was estimated to be lower for those in teaching hospitals compared with non-teaching hospitals [40]. One study estimated that private insurance paid 37% more than traditional Medicare, with Medicare Advantage paying an additional 10% above this, for inpatient care for five common inpatient admissions, even after controlling for enrollee and hospital mix [41]. Chernew et al. found, similarly, that commercial fees were about double that of Medicare in an examination of state-level price variation [42]. Health and economic outcomes may also vary by physician. For example, being assigned to an experienced general surgeon was found to reduce the risk of severe complications and mortality by 5% versus being assigned to a newly graduated surgeon [3]. Non-patient preferences, such as those held by the public and healthcare providers are contained in this class and may be relevant to consider depending on the decision problem that the economic evaluation is intended to inform. For example, one economic analysis of the value of intensive versus conventional glucose control with oral drugs in different age intervals of older adults living with diabetes found that, when US population-based utility weights for health states and treatments were used, ICERs ranged from $136,000 for those aged 65–70 years to $24 million for those aged 90–95 years [43]. However, when utility weights derived from these patients were applied, the ICERs were dramatically lower, by 51% ($67,000) for those aged 60–65 years and 94% ($134,000) for those 90–95 years, respectively. In addition, unlike single-payer systems, drug-market intermediaries—like pharmacy benefit managers who negotiate with manufacturers on behalf of insurers, and employers, many of whom sponsor health insurance and manage employee benefits play a significant role in the US healthcare system [44].

In contrast to the first five classes, which consist of factors observable at the time of the treatment decision, the final class —"factors revealed *ex post*"—includes factors that are only known after treatment exposure. The most intuitive example is the response to treatment. For instance, the continuation of weight-loss treatments often depends on achieving a minimum weight reduction (e.g., 5%) [45, 46]. Similarly, an economic evaluation that incorporated intensive guideline-directed medical therapy optimization

for patients hospitalized with heart failure found that continued assessment of dose and patient adherence was cost effective versus usual care in the USA [47]. The occurrence of adverse events, which could potentially be avoided by adjusting dose or switching treatments, is another factor in this class. In the US setting, this category can also include costs and patient co-payments that remain uncertain until claims are adjudicated, as differences in these can influence the probability of treatment continuation, and hence health outcomes.

2.2 Including Heterogeneity in a Comparative Economic Analysis

There is a large body of methods to address heterogeneity in a comparative economic analysis. Much of this work is summarized in three, mostly non-overlapping, literature reviews that serve as useful references [48–50]. The work steps can loosely be divided into design, analysis, and reporting phases. We consolidate key points from the literature reviews and provide additional context below, but we refer the reader to the previous reviews and to the original sources for full details.

2.2.1 Design Phase

The possibility that patient and non-patient heterogeneity may influence health and outcomes should be considered in most, if not all, comparative economic analyses. Identification of what sources may be relevant, if any, ideally begins in the planning stage of a study. This process should follow the scientific method, utilizing evidence review, testable hypotheses, and analysis [12]. Given the complexity of heterogeneity in healthcare, a multidisciplinary team is crucial to ensure all perspectives are considered. This could involve experts from fields like sociology, psychology, and health informatics, among others. Patient input is also essential to align the analysis with the needs of impacted populations [51]. Pre-specifying subgroups can enhance study credibility, but it may be infeasible—particularly early in a treatment's life cycle—because of limited data and knowledge [2, 12].

A number of considerations should be taken into account when designing the study: (1) which sources of heterogeneity to address, (2) evidence availability and opportunities for generating new evidence, (3) suitable statistical methodologies and analytical frameworks, (4) ethical implications of stratified decision making, and (5) practical barriers or challenges.

Each of the six classes of heterogeneity outlined in Sect. 2.1 should be considered for inclusion. While comprehensive, these classes are not mutually exclusive. Multiple factors may be relevant for a given application and any individual factor may affect value through multiple pathways, for example, baseline LDL levels can influence both relative and absolute treatment effects.

Selection of which sources of heterogeneity to address should be evidence driven. In addition to clinical trial data, surveys, systematic literature reviews, meta-analyses, expert panels, and new analyses of existing data can provide important information. In the USA, RWD play an important role as market access decisions are frequently updated; insurers routinely revise formularies (e.g., annually) and open enrollment periods allow insured populations to switch plans. Where generating new quantitative evidence is impractical, a mixed-methods approach—combining qualitative and quantitative analyses—can be used to synthesize measurable and unmeasurable factors in a cohesive decision-making framework [52]. Shields et al., for example, recommend using logic models to structure theories, evidence, expert opinion, assumptions, and preferences [12], an approach recently adopted by the US Food and Drug Administration in a draft guidance aimed at standardizing risk management programs [53].

The ethical implications of stratified resource allocation decisions should be carefully considered, ideally with input from a multidisciplinary team and patients with direct disease experience [1, 2, 12]. Even when economically efficient, decisions that deny care—particularly when based on characteristics like age, sex, disability, and race—may violate fairness norms. Stratified decisions may be defensible when clear evidence shows differing treatment effects across subgroups [1, 2, 12]. For example, older patients may have a diminished immune response so that the efficacy of immunotherapies is severely limited [54]. Stratified decision making does not always exacerbate disparities; in some cases, they can serve to reduce them [1]. When stratified decision making is deemed unethical, a subgroup analysis can nonetheless help inform decision makers as to the opportunity costs of a population level decision [1, 2, 48].

Finally, potential barriers to implementing stratified decisions in routine practice should be examined. These include challenges in observing relevant factors or identifying appropriate proxies, the costs of gathering necessary data (e.g., biomarker testing), and insufficient data availability (particularly for rare diseases) [1, 2, 48]. Additional concerns may arise if subgroup membership is based on information that could be gamed (e.g., a weight threshold to qualify for weight loss therapy) [2].

2.2.2 Analysis Phase

In the analysis phase, study methods are applied to study data to determine whether and how heterogeneity should be parameterized in the comparative economic analysis. Statistical methods—including an exploratory analysis, hypothesis

testing, and a confirmatory analysis—can help to determine whether heterogeneity should be explicitly addressed, which types of heterogeneity should be considered, the optimal subgroup definitions, and the empirical estimates of these effects to be incorporated into the analysis. These methods must be aligned with the data that are available.

Willke et al. reviewed eight statistical methods for addressing heterogeneity of treatment effects, which is a factor in Class 1 [49]. While the intended readership was researchers interested in incorporating heterogeneity of treatment effects into randomized controlled trials, the paper also discusses issues related to non-randomized data. The eight methods covered include: conventional subgroup analysis of a clinical trial, subgroup analysis of meta-analyses and meta-regression, predictive risk modeling, classification and regression tree analysis, latent growth modeling/growth mixture modeling, series of *n of 1* trials (i.e., repeated cross-over trials for single patients), quantile treatment effect regression, and non-parametric methods like kernel smoothing methods. The authors provide a description for each method.

The trade-off between increasing subgroup specificity and greater uncertainty is inherent to all eight methods, as dividing the population into smaller groups—whether through predictive modeling, a subgroup analysis, or non-parametric approaches—often leads to higher variability in estimates and reduced precision, especially when sample sizes within subgroups become smaller. Testing multiple subgroups increases the likelihood of detecting spurious but statistically significant associations by chance, but this risk may be mitigated with statistical adjustments to control type I error, such as the Bonferroni correction.

Conventional methods such as a subgroup analysis and meta-analysis typically generate readily interpretable parameters, such as treatment effects, that can be directly incorporated into comparative economic models. In contrast, methods like classification and regression tree, latent growth modeling/growth mixture modeling, and non-parametric approaches may not directly align with the parameters needed for economic modeling and may require additional steps to derive model-ready parameters, such as treatment effects. Non-parametric methods, which avoid assumptions about data distribution, may also require transformation of key outputs, such as distributions of costs or effects, to generate parameters suitable for integration into economic models. Individual methods are mapped to suitable applications based on key data considerations—including the types of evidence available (e.g., theory, trial evidence from various phases, and RWD), the purpose of the analysis (e.g., exploration, testing, or confirmation), the character of the data (e.g., randomized controlled trial vs RWD), and the volume of data. For example, a meta-analysis, which pools smaller studies into a larger sample, can enhance statistical power but is only useful when the studies to be included are sufficiently homogeneous.

Grutters et al. [48] and Shields et al. [50] surveyed methods specifically used to address patient heterogeneity in the economic evaluation literature, with the former covering methods up to 2011 and the latter updating this for the period from 2011 to January 2024. Prior to 2011, this was generally limited to routine subgroup analyses and to regression methods that link baseline risk, treatment effects, utility, and resource utilization to patient characteristics such as demographics, clinical factors, and preferences. For instance, ordinary least squares regression of net monetary benefit on patient characteristics and interaction terms has been used to identify relevant subgroups [55]. Building on this, the same authors introduced the use of a system of seemingly unrelated equations regression, which relaxes the structural constraints of ordinary least squares by allowing for separate equations with different functional forms and covariates, thereby enhancing statistical efficiency [56]. Nixon and Thompson adapted the seemingly unrelated equations framework by applying Bayesian methods, which allowed for the incorporation of informative priors and extended its applicability (e.g., to non-randomized studies) [57]. This could be particularly valuable in settings with small sample sizes, as it provides more robust estimates by incorporating both data-driven information and expert judgment. When aggregate-level data are used to inform regressions (i.e., meta-regression), researchers should consider the risk of ecological fallacy (i.e., a faulty assumption that relationships observed at the group level necessarily apply to individuals within those groups) [12].

Shields et al. updated this review to cover the period from 2011 to January 2024 [50], identifying six additional methodologies for addressing patient heterogeneity in economic evaluations. Machine learning techniques, such as causal forests, analyze heterogeneity by estimating causal treatment effects across subgroups of patients. Causal forests build a large number of decision trees to identify variation in treatment effects and outcomes, including costs and net monetary benefits, within different patient characteristics, offering a data-driven non-parametric approach to detect these effects in a more automated and scalable manner [58]. Local instrumental variables—variables that are correlated with the treatment but not directly with the outcome—have been used to address confounding and heterogeneity in observational data. Local instrumental variables help reduce biases from unobserved confounding factors, which could otherwise distort estimates of treatment effects [59]. Subpopulation Treatment Effect Pattern Plot techniques provide a graphical exploration of treatment effect variations across different subpopulations [60]. The Subpopulation Treatment Effect Pattern Plot makes no assumptions about the nature of the

relationship between outcomes and covariates within each treatment group, allowing for a flexible data-driven approach to identify subgroups where treatment is most effective. Multistate statistical modeling, whereby model-based transition probabilities are derived to account for baseline patient characteristics, is well suited for populating microsimulation models [50]. Use of regression-based approaches to estimate preference heterogeneity depends on the data available [61]. When individual-level utility data are available, as opposed to aggregate health state data, regression can be used to directly model the effects of individual characteristics on outcomes. Patient preference heterogeneity has been assessed using discrete choice experiments, which involve presenting individuals with hypothetical treatment options to capture their preferences for various treatment attributes [37].

No studies, to our knowledge, have surveyed methods to address sources of non-patient heterogeneity, such as surgeon experience, care provider setting, and insurance levels, but there are empirical studies that have considered factors in this class. For example, McClellan et al. used differences in distance to a catheterization hospital—an observable factor that influences treatment decisions but does not directly affect patient outcomes—as an instrumental variable to address unobserved confounding related to underlying patient health, enabling unbiased estimates of the short- and long-term mortality effects of intensive treatments for acute myocardial infarctions [29]. Addressing heterogeneity unrelated to patients requires data structured around these non-patient factors, such as individual healthcare providers or facilities. When such non-patient data are available in randomized controlled trials or RWD, it can be reorganized to explore how variations in these factors influence outcomes. By categorizing or stratifying outcomes based on these factors, statistical techniques like hierarchical or multi-level modeling can be used to quantify the impact of non-patient heterogeneity. In other cases, it may be necessary to collect new data, if feasible and/or use a mixed-methods approach to leverage both qualitative and quantitative data.

For cases where empirical estimates can be generated for subgroups, it is important to examine not only statistical significance, but also clinical and economic meaningfulness. The magnitude of the statistically significant difference should be large enough to justify changes in clinical practice or policy [2]. Replication of subgroup results across multiple datasets may help reduce the risk of spurious findings and increase credibility [62]. Whether pre-specified or not, subgroup selection should be justified, with careful consideration of plausibility.

2.2.3 Implementation in Economic Models

Once the decision is made to address specific sources of heterogeneity, the next step is incorporating these factors into the comparative economic analysis. How to parameterize heterogeneity will depend on whether the analysis will be performed directly using patient-level data, in which case heterogeneity is captured naturally, or whether economic simulation modeling is used, in which case it must be explicitly incorporated. Note that the extent of observable heterogeneity may be more limited with clinical trial data than with RWD, owing to study features like exclusion criteria and treatment protocols.

Grutters et al. identified several methods for incorporating systematic relationships into economic models in use before 2011 [48]. Early modeling studies often adapted cohort models to account for heterogeneity by subdividing populations into subgroups, sometimes using tunnel states to reflect patient history. However, these adaptations can lead to a rapid increase in the number of health states and parameters, making models computationally challenging and harder to manage. Patient-level microsimulation methods, such as semi-parametric Markov modeling and discrete event simulation, are more flexible alternatives. These methods use tracker variables to tailor event risks, costs, and utilities to individual characteristics, providing a more granular approach to modeling heterogeneity [63]. They are also useful when factors revealed *ex post* are important. Methods for incorporating heterogeneity into economic models include using subgroup-specific parameters (e.g., age-specific mortality rates, costs, and utility values) [64–67], as well as regression equations that link patient characteristics to outcomes such as event probabilities, survival rates, and utility values.

The Shields et al. review identified two additional modeling methods that were found first in the literature post-2011 [50]. Murphy et al. used a Bayesian hierarchical modeling approach to identify heterogeneity in clinical response by tumor histology and estimate the probability of response for each histology [69]. Vemer et al. highlighted the importance of distinguishing heterogeneity from parameter uncertainty by using double-loop PSA [70].

The methods described here cannot alone determine the optimal strata for decision making, quantify the value of stratified decisions—or, conversely, the opportunity cost of not making them—or assess the potential net benefit of additional evidence collection to reduce decision-making uncertainty. Five analytical frameworks have been proposed

to perform such analyses [22, 71–74], with a 2022 literature review summarizing the four frameworks that were available at that time [11]. When stratified results are not reported for ethical reasons, an equity-efficiency impact analysis using these frameworks has been proposed [1, 2, 10, 12]. Briefly, these frameworks maximize net benefits (e.g., net monetary benefits [NMB]) by determining the optimal treatment strategy for subgroups or individuals in heterogeneous populations and estimate the value of individualized or stratified decisions versus population-wide decisions. They primarily differ in how they address parameter uncertainty, identify and define subgroup stratifications (e.g., including perfect and imperfect information, as well as variability in treatment adoption), and what metrics are considered [11]. The context of the study—including the nature of the intervention, the availability of subgroup-level treatment effect and uptake data, and the potential for additional evidence collection—should guide the selection of an appropriate analytical framework [11].

The "Value of Stratification" framework focuses on maximizing NMB by ensuring interventions are used only in cost-effective subgroups, introducing a novel metric for losses because of non-adherence [71]. The primary outcome is the net benefit from stratification, calculated as the difference between total NMB for cost-effective subgroups and total NMB across all subgroups. While the authors did not explicitly address parameter uncertainty, probabilistic modeling can be used with Value of Stratification to account for it. The "Expected Value of Individualized Care" extends stratification to the individual level, quantifying the full cost of ignoring patient-level heterogeneity and introducing a novel metric for loss due to imperfect personalization [72, 75]. The cost of ignoring heterogeneity is measured as the difference between NMB from fully individualized care and NMB from population-based care. The "Value of Heterogeneity" framework, building on Value of Stratification and Value of Stratification, estimates optimal subgroup stratifications, with the existing evidence base in the static form, and with an enhanced evidence base in the dynamic form [73, 76]. Value of information statistical methods, which quantify the value of reducing uncertainty by collecting additional data, are used to define optimal stratifications in a dynamic analysis. All possible combinations of heterogeneity factors are considered, and an efficiency frontier is defined. The "Loss with Respect to Efficient Diffusion" framework differs from the others by focusing on the realized value of technologies, accounting for heterogeneity in treatment uptake patterns [22]. Efficient diffusion occurs when a treatment is adopted only in cost-effective subgroups, and losses from inefficient adoption are quantified as the difference in NMB between efficient and inefficient adoption. Finally, the "Choice Set" framework considers sets of treatments under alternative insurance policies, incorporating real-world features like endogenous treatment selection, heterogeneous treatment uptake, and patient preferences [74].

2.2.4 Reporting Phase

Transparent and comprehensive study reporting enhances reproducibility and facilitates correct interpretation and critical appraisal of study findings [12]. Specifically, subgroup selection criteria and analytical methods should be documented, as should a clear rationale and empirical justification for both including and excluding subgroups [12]. When a subgroup analysis has been performed, the full set of results should be presented and plausible explanations should be provided for differences that may occur across subgroups [68]. When a subgroup analysis is not performed, the probable implications should be discussed [2, 12]. If a formal equity-efficiency trade-off analysis is carried out, the results should be presented as well. All limitations and assumptions should be acknowledged.

3 Empirical Practices Related to Heterogeneity in the USA

Despite its critical importance for making efficient and equitable resource allocation decisions, relatively little is known about how well researchers address heterogeneity in a comparative economic analysis. Most of the existing knowledge comes from a single study that examined a sample of 200 cost-utility studies published in 2014, drawn at random from the Tufts Medical Center Cost-Effectiveness Analysis Registry[2] [14]. First, Lavelle et al. found that a subgroup analysis was infrequently reported, appearing in only 19% of the full sample and 25% of the 68 studies in the US subsample. Second, they found that subgroup analyses were primarily limited to a few stratifying variables: age was used in 35 of the 68 studies with a subgroup analysis, while clinical factors (e.g., laboratory results, tests, or procedures) and a history of comorbidities were each used in seven studies, and predicted risk was used in six studies. Among the 35 studies reporting only one source of variability, baseline risk was the most commonly varied input variable (16 studies), followed by treatment effects (nine studies), life expectancy (seven studies), and costs (three studies). Notably, no studies were stratified by preferences. Third, they observed that cost effectiveness for subgroups frequently

[2] The Tufts Cost-Effectiveness Analysis Registry is a comprehensive database containing detailed information on over 10,000 cost-utility analyses that quantify health benefits in terms of quality-adjusted life-years or disability-adjusted life-years.

differed from that of the overall population in the US subsample. Specifically, 57% of the reported stratifications contained at least one subgroup with a value determination that differed from that of the overall population (e.g., cost effective for a subgroup but not for the population, or vice versa). To address questions unanswered by Lavelle et al. [14], we retrieved the sample of 68 US studies, hereafter referred to as the Lavelle Sample,[3] and performed an additional analysis.

First, we explored the question of why a subgroup analysis was absent in 80% of the sample studies. A valid reason for the omission could be insufficient heterogeneity for that study problem to a warrant subgroup analysis, confirmed by careful analysis. Supporting this, value determinations for subgroups and the overall population were consistent for 43% of the stratifications analyzed by Lavelle et al. [14]. Ethical concerns, such as the risk that stratified decisions might have disadvantaged vulnerable populations [1, 2, 10, 15], might have also justified the omission of a subgroup analysis. Lack of suitable data, especially for rare conditions with limited patient populations, might have been a reason for certain studies. Transparent reporting of the rationales for both including and excluding a subgroup analysis and a discussion of the potential implications is recommended practice [12], so we searched these studies for confirmation. None of the 49 studies that omitted a subgroup analysis provided an explicit rationale, nor did any of the 19 studies with subgroup-level results for not considering additional sources. Only two studies with subgroup-level results provided a rationale for its inclusion [79, 80]. Given the dearth of studies explaining why or why not a subgroup analysis was performed, we could not determine whether the absence of a subgroup analysis in 80% of the sample is warranted.[4]

Second, we used descriptions of heterogeneity in the original articles to impute classes considered using the informal taxonomy described in Sect. 2.1. The results are presented in column 1 in Table 2. Factors related to the intervention-related, Class 1, and factors unrelated to both the intervention and disease, Class 3, were each considered in roughly one third of the studies. Disease- but not intervention-related factors, Class 2, were used to stratify subgroups in 84% of the sample. None of the studies accounted for heterogeneity in patient preferences, non-patient factors, or sources of heterogeneity unknowable *ex ante*, but known *ex post*.

These results reflect empirical practices over a decade old and may not reflect research practices since that time. We searched for more recent evidence on the use of a subgroup analysis in a comparative economic analysis in several ways[5]; however, no additional studies were found.To address this gap, we created a sample of US studies published from 2015 to 2022 using the Tufts Medical Center Cost-Effectiveness Analysis Registry terms: "subgroup," "sub-group," "stratify," "stratified," stratification," or "subset," which returned 106 records. Eighty-five of these studies included subgroup-level results, which are hereafter referred to as IHE Sample [see the Electronic Supplementary Material [ESM] for a full bibliography). While one study reported a subgroup analysis for seven different stratification variables—age, background treatment, sex, predicted risk, smoking status, race, and severity [81] —74% of the studies reported results for just one subgroup stratification. Age and biomarkers were the most common individual characteristics by which subgroups were stratified, 38% and 30%, respectively (see Table 2 of the ESM). The percentage of studies including each class of heterogeneity are presented in column 2 of Table 2. Disease- but not intervention-related factors (Class 2) were the most common and were present in 89% of the studies. Intervention-related factors (Class 1) and factors unrelated to both intervention and disease (Class 3) were addressed in 15% and 18% of the studies, respectively. No studies considered patient preferences (Class 4) or factors revealed *ex post* (Class 6). Non-patient factors (Class 5) were included in two studies (2%). They included: residence in a long-term facility versus in the community [82], and state of residence as a proxy for heart transplant donor

[3] We reclassified two studies, [77, 78] which were found to report subgroup analysis, increasing the share of studies addressing heterogeneity from 25% to 28%.

[4] Two alternative explanations, which were not testable with this sample, might have impacted our results. First, as the Tufts Medical Center Cost-Effectiveness Analysis Registry is restricted to a cost-utility analysis, almost all of the studies examined were performed with economic modeling. A comparative economic analysis, performed directly with patient-level RWD, may have a natural advantage in supporting a subgroup analysis as the participants in clinical trials must meet inclusion criteria. Second, subgroup analyses can be reported as separate publications, which may not be captured in random sampling of the literature.

[5] First, we identified all studies that referenced Lavelle et al. [14] and reviewed them for relevance, as authors of similar, but newer, studies would likely be aware of this seminal study. Second, we used the "Similar Articles" PubMed tool to identify additional relevant studies, occurring both prior to and after publication of Lavelle et al. [14]. Third, we reviewed the reference list in Lavelle et al. [14] as well as references in key methodological studies (Sculpher, [2] Kohli-Lynch and Briggs, [1] Willke et al., [49] Grutters et al., [48] and Shields et al. [12, 50]). Fourth, we reviewed the results from a May 2023 PubMed search with search terms "patient heterogeneity" AND "cost-effectiveness".

Table 2 Percent of studies including each class of heterogeneity, by study sample

Heterogeneity class	Lavelle sample ($n = 19$) (%)	IHE Sample ($n = 85$) (%)
(1) Intervention-related factors	32	15
(2) Disease- but not intervention-related factors	84	89
(3) Factors unrelated to both intervention and disease	32	18
(4) Patient preferences	0	0
(5) Non-patient factors	0	2
(6) Factors revealed *ex post*	0	0

Column totals do not sum to 100% as a study can capture more than one class of heterogeneity

waiting times [83]. The distribution was, for the most part, similar to that in the Lavelle Sample (see Table 2).

Third, we updated the value determination analysis using the 16 US studies reporting results for both the overall population and individual subgroups in the Lavelle Sample. We found that at least one subgroup was cost effective when the overall population was not, or vice versa, for 48% of the 27 subgroup stratifications (i.e., the factors upon which the subgroupings are based on). Some studies reported multiple stratifications (e.g., one study had separate factor stratifications for age, sex, and smoking status [13]) [see Table 1 of the ESM]. A limitation of this analysis is that changes in value determination are sensitive to the cost-effectiveness threshold used. Minor changes in ICERs can lead to reversals in value assessments when estimates are close to the threshold. To address this, we calculated relative differences between ICERs for the overall population and each subgroup. The frequency distribution is presented in Fig. 1. The *y*-axis depicts the percent of subgroups whose ICERs deviate from the overall population ICER by amounts that ranged from > 10 to > 100% (*x*-axis). Roughly one third of subgroups had ICERs that deviated from the overall population ICER by at least 50%. Only two studies provided explanations for divergent results [84, 85]. Among the 59 studies in the IHE Sample that reported results for at least one subgroup and for the overall population, roughly one third of the subgroups had ICERs that diverged by at least 50% from the overall average estimate and nearly 10% diverged by at least 100%.

4 Discussion

Decision making in the decentralized US healthcare system contrasts sharply with that of most other industrialized countries, many of which are characterized by a dominant single payer. This study aimed to enhance understanding and promote the application of heterogeneity in US-focused economic evaluations. The implications for healthcare decision making are significant: neglecting to account for the various sources of heterogeneity that can influence outcomes may lead to suboptimal market access decisions, missed opportunities for personalized interventions, and reduced economic efficiency in resource allocation. To achieve these objectives, we built upon previous work describing heterogeneity, synthesized earlier reviews of methods, and generated new evidence on how, and investigated why or why not, heterogeneity was incorporated in empirical research.

We found that Sculpher's informal taxonomy provides a good general description of relevant types of heterogeneity [2], even in the US context. We suggested several modifications to reduce the overlap between classes and emphasize the importance of both patient and non-patient preferences. To highlight the relevance of non-patient factors and preferences that are key determinants of costs and outcomes in the US setting, we provided specific examples. The informal taxonomy presented in our paper can serve as a useful checklist for researchers in the field. However, future work to develop a formal taxonomy—complete with mutually exclusive classes and methodological recommendations tailored to each class—would help to standardize empirical methods around best practices.[6] The topic of how to deal with heterogeneity in economic evaluations is central to ISPOR's Science Strategy [86], and the organization is in the process of convening a Task Force to develop guidelines. Leveraging the collective expertise of a broad group of stakeholders to refine the taxonomy and establish tailored methodological approaches for each category would represent two invaluable contributions to the field.

While a wide range of methods suitable for addressing the many types of heterogeneity exist and have been summarized [48–50], numerous factors may have hindered their implementation. These include challenges with prespecifying subgroups, resource and data requirements, and ethical concerns about stratified decisions potentially disadvantaging vulnerable populations. Data availability can

[6] A formal taxonomy requires mutually exclusive categories. While this informal taxonomy adheres to this principle in some cases — for example, a factor cannot be both knowable both *ex ante* and *ex*

be a particularly significant hurdle, as high-quality granular data are often required to perform a subgroup analysis effectively. For instance, powering registrational clinical trials to explore treatment effect heterogeneity can be infeasible and/or significantly increase costs. As a result, analyzing small subgroups from an underpowered trial gives rise to a trade-off that researchers need to consider, namely between the added uncertainty associated with stratification and the bias introduced from using unstratified results. The Value of Heterogeneity analytical framework is designed to assess this trade-off and estimate the net value of collecting additional evidence to reduce uncertainty [73, 76]. Ramaekers et al. [87] noted that the vagueness of methodological directives in Health Technology Assessment (HTA) guidelines may have contributed to inconsistent or suboptimal use of a subgroup analysis, inadvertently downplaying the importance of addressing heterogeneity. These barriers are compounded by ethical concerns regarding the fairness of stratified decisions, especially when certain subgroups, such as vulnerable populations, may be disadvantaged. A critical examination of how researchers balance these trade-offs would be valuable. It should be noted that while we discuss many methods, we did not perform a systematic review, but rather leveraged on four previous reviews [11, 48–50]. Moreover, as our aim was to explore the range of methods available, researchers interested in applying a method should consult the original manuscript for full details.

We found just one study that examined the extent to which comparative economic analyses account for heterogeneity through a subgroup analysis, with roughly one in four US studies addressing this issue [14]. Value determination reversals were common in the studies that reported cost-effectiveness results for both subgroups and the overall population. However, this finding could have been impacted by selection bias (i.e., studies showing small variations in cost effectiveness are less likely to report a subgroup analysis) and by the relatively small sample size. Surprisingly, none of the sample studies provided an explicit rationale for excluding a subgroup analysis and only two studies provided a rationale for including it. As such, it was impossible to determine whether omission of a subgroup analysis was reasonable. Many of the studies reporting a subgroup analysis focused solely on a few stratification factors, with age and clinical factors being the most common. Notably, none of these studies addressed non-patient sources of heterogeneity or patient preference heterogeneity. We identified two studies that examined non-patient heterogeneity in a more recent sample of studies published between 2015 and 2023, but the overall extent to which different types of heterogeneity were considered in cost-utility analyses was largely unchanged. Although this sample was larger, the search terms specified may not have identified all relevant studies. Note that these findings are limited to cost-utility studies as we relied on data from the Tufts Medical Center Cost-Effectiveness Analysis Registry and may not generalize to other forms of comparative economic analyses. An area for future research is the extent to which, and how, US real-world comparative economic studies address heterogeneity.

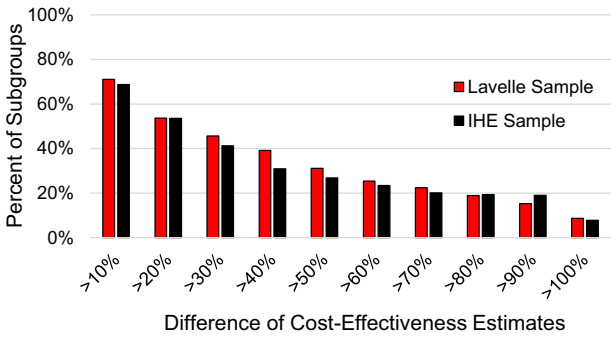

Fig. 1 Relative difference between subgroup-level and overall population cost-effectiveness estimates

Although the evidence review could not determine whether a subgroup analysis was optimally considered given the absence of rationales for inclusion or exclusion, improved data and clearer guidance may help to promote the quality and consistency of addressing heterogeneity in economic evaluations. For evaluations that extrapolate clinical trial results with modeling, researchers should consider investigating heterogeneity in RWD from electronic health records and claims databases—the latter of which can even provide information on key non-patient factors in the US setting like insurance coverage. The evidence generated from RWD can complement clinical trial data, enabling a better subgroup analysis and may help to alleviate statistical issues related to underpowered trials. Mixed methods, including those that combine quantitative information with qualitative insights from patients and other stakeholders, may be useful when considering patient and non-patient preferences as well as equity issues. In addition, greater use of analytical frameworks that estimate the value of stratification and guide

Footnote 6 (continued)

post, and a patient-related factor cannot simultaneously be a non-patient-related factor — there are notable overlaps among some classes. Consider Classes 1–3, which segment factors based on two binary criteria: whether they are intervention related or not, and disease related or not. This segmentation logically produces four, rather than three, possible combinations. Moreover, a factor can simultaneously influence these four classes and patient preferences, leading to eight potential combinations. For instance, "fear of needles" in a comparison of inhaled versus injectable insulin for diabetes illustrates this overlap. In patients with needle phobia, preferences for inhaled insulin would likely be stronger. At the same time, the relative benefit of inhaled insulin over injectable insulin may be greater for this subgroup compared with others, potentially mediated through differences in adherence.

data collection could help decision makers better evaluate the trade-offs between the uncertainty of stratification and the bias of unstratified results. As highlighted earlier, creating concrete recommendations tailored to different types of heterogeneity that consider health equity implications and promoting greater consistency in empirical practices would be important contributions of the ISPOR Task Force on heterogeneity.

5 Conclusions

Despite well-documented methods to address heterogeneity, its application is limited in US cost-utility studies, especially regarding patient preferences and non-patient factors. As these factors often drive real-world health outcomes and costs in the USA, proper consideration of, and reporting on heterogeneity is essential to avoid erroneous market access decisions, suboptimal patient outcomes, and economic inefficiency. In cases where addressing heterogeneity is deemed inappropriate, such as for ethical reasons, researchers should clearly explain these decisions and their potential implications. These findings should serve to caution decision makers who utilize value assessment information and underscore the need for improved practices. Future efforts should include multidisciplinary competence and importantly the patient perspective. In addition, work to refine the heterogeneity taxonomy and promote best methods is warranted and should be considered by the planned ISPOR Task Force on heterogeneity.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40273-025-01478-z.

Acknowledgments The authors thank the three anonymous reviewers, whose valuable comments served to enhance this article. Editorial assistance was provided by Panita Maturavongsadit, PhD, of Lumanity Communications Inc., and was funded by Janssen Scientific Affairs, LLC, a Johnson & Johnson Company.

Declarations

Funding This study was funded by Janssen Scientific Affairs, LLC, a Johnson & Johnson Company. The authors had full control of the article and the findings were not contingent on sponsor approval.

Conflicts of Interest/Competing Interests Michael S. Willis and Andreas Nilsson are employees of the Swedish Institute for Health Economics, which provides consulting services for governmental bodies, academic institutions, and commercial life science enterprises. Cheryl A. Neslusan is an employee of Janssen Scientific Affairs, LLC, a Johnson & Johnson Company.

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Availability of Data and Material Full extraction of data is available upon reasonable request.

Code Availability Not applicable.

Authors' Contributions MW acts as the overall guarantor of the article. All authors were involved in the concept and design of the study. Analysis was performed by MW and AN. All authors interpreted the results. Drafting of the initial manuscript was performed by MW and AN and all authors provided revisions for the final draft. All authors confirm that they meet the International Committee of Medical Journal Editors uniform requirements for authorship and that they have contributed to the critical analysis and interpretation of the data and drafting and/or critically revising the article and share the final responsibility for the content of the manuscript and the decision to submit it for publication.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc/4.0/.

References

1. Kohli-Lynch CN, Briggs AH. Heterogeneity in cost-effectiveness analysis. Oxford: Oxford University Press; 2019.
2. Sculpher M. Subgroups and heterogeneity in cost-effectiveness analysis. Pharmacoeconomics. 2008;26(9):799–806.
3. Howard RA, Thelen AE, Chen X, Gates R, Krumm AE, Millis MA, et al. Mortality and severe complications among newly graduated surgeons in the United States. Ann Surg. 2024;279(4):555–60.
4. Barbieri M, Drummond M, Willke R, Chancellor J, Jolain B, Towse A. Variability of cost-effectiveness estimates for pharmaceuticals in Western Europe: lessons for inferring generalizability. Value Health. 2005;8(1):10–23.
5. Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, et al. Generalisability in economic evaluation studies in healthcare: a review and case studies. Health Technol Assess. 2004;8(49):1–192 (**iii–iv**).
6. Lin CC, Hirai AH, Li R, Kuklina EV, Fisher SK. Rural-urban differences in delivery hospitalization costs by severe maternal morbidity status. Ann Intern Med. 2020;173(11 Suppl.):S59-62.
7. Palmer AY, Wade AJ, Draper B, Howell J, Doyle JS, Petrie D, et al. A cost-effectiveness analysis of primary versus hospital-based specialist care for direct acting antiviral hepatitis C treatment. Int J Drug Policy. 2020;76: 102633.
8. Phelps CE. Good technologies gone bad: how and why the cost-effectiveness of a medical intervention changes for different populations. Med Decis Making. 1997;17(1):107–17.
9. National Pharmaceutical Council. The myth of average: why individual patient differences matter. 2022. https://www.npc-now.org/sites/default/files/2022-01/The_Myth_of_Average_01.2022.pdf. Accessed Mar 2024.

10. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second Panel on Cost-Effectiveness in Health and Medicine. JAMA. 2016;316(10):1093–103.
11. Pataky RE, Bryan S, Sadatsafavi M, Peacock S, Regier DA. Tools for the economic evaluation of precision medicine: a scoping review of frameworks for valuing heterogeneity-informed decisions. Pharmacoeconomics. 2022;40(10):931–41.
12. Shields GE, Wilberforce M, Clarkson P, Farragher T, Verma A, Davies LM. Factors limiting subgroup analysis in cost-effectiveness analysis and a call for transparency. Pharmacoeconomics. 2022;40(2):149–56.
13. Black WC, Gareen IF, Soneji SS, Sicks JD, Keeler EB, Aberle DR, et al. Cost-effectiveness of CT screening in the National Lung Screening Trial. N Engl J Med. 2014;371(19):1793–802.
14. Lavelle TA, Kent DM, Lundquist CM, Thorat T, Cohen JT, Wong JB, et al. Patient variability seldom assessed in cost-effectiveness studies. Med Decis Making. 2018;38(4):487–94.
15. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. JAMA. 1996;276(14):1172–7.
16. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. JAMA. 1996;276(16):1339–41.
17. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. JAMA. 1996;276(15):1253–8.
18. American Diabetes Association (ADA). Standards of care in diabetes: 2023. Diabetes Care. 2023;46:S1–291.
19. Navarese EP, Andreotti F, Raggi P, Kolodziejczak M, Buffon A, Bliden K, et al. Baseline low-density lipoprotein cholesterol to predict the extent of cardiovascular benefit from lipid-lowering therapies: a review. Eur Heart J Cardiovasc Pharmacother. 2019;5(1):47–54.
20. Wilder ME, Kulie P, Jensen C, Levett P, Blanchard J, Dominguez LW, et al. The impact of social determinants of health on medication adherence: a systematic review and meta-analysis. J Gen Intern Med. 2021;36(5):1359–70.
21. Whitman A, De Lew N, Chappel A, Aysola V, Zuckerman R, Sommers B. Addressing social determinants of health: examples of successful evidence-based strategies and current federal efforts. 2022. https://aspe.hhs.gov/sites/default/files/documents/6ba4bbb2e9c9551355a6926f023f1585/SDOH-Evidence-Review.pdf. Accessed 16 Dec 2024.
22. Kim DD, Basu A. New metrics for economic evaluation in the presence of heterogeneity: focusing on evaluating policy alternatives rather than treatment qlternatives. Med Decis Making. 2017;37(8):930–41.
23. Mora PF, Sutton DR, Gore A, Baliga B, Goldfaden RF, Nikkel C, et al. Efficacy, safety and cost-effectiveness comparison between U-100 human regular insulin and rapid acting insulin when delivered by V-Go wearable insulin delivery device in type 2 diabetes. BMJ Open Diabetes Res Care. 2020;8(2):e001832.
24. Edmonds T, Roberts G, Cheng J, Mumford A. HTA39 estimating the true cost of weight-based chemotherapy. Value Health. 2022;25(12):S303–4.
25. Mihaylova B, Emberson J, Blackwell L, Keech A, Simes J, Barnes EH, et al. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. Lancet. 2012;380(9841):581–90.
26. Mihaylova B, Wu R, Zhou J, Williams C, Schlackow I, Emberson J, et al. Lifetime effects and cost-effectiveness of standard and higher-intensity statin therapy across population categories in the UK: a microsimulation modelling study. Lancet Reg Health Eur. 2024;40: 100887.
27. Morton JI, Marquina C, Lloyd M, Watts GF, Zoungas S, Liew D, et al. Lipid-lowering strategies for primary prevention of coronary heart disease in the UK: a cost-effectiveness analysis. Pharmacoeconomics. 2024;42(1):91–107.
28. Briggs A, Mihaylova B, Sculpher M, Hall A, Wolstenholme J, Simoons M, et al. Cost effectiveness of perindopril in reducing cardiovascular events in patients with stable coronary artery disease using data from the EUROPA study. Heart. 2007;93(9):1081–6.
29. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. JAMA. 1994;272(11):859–66.
30. Kaplan RM, Hays RD. Health-related quality of life measurement in public health. Annu Rev Public Health. 2022;5(43):355–73.
31. Heijdra Suasnabar JM, Finch AP, Mulhern B, van den Akker-van Marle ME. Exploring the measurement of health related quality of life and broader instruments: a dimensionality analysis. Soc Sci Med. 2024;346: 116720.
32. Phelps CE, Lakdawalla DN. Valuing health: the generalized and risk-adjusted cost-effectiveness (GRACE) model. Oxford: Oxford University Press; 2024.
33. Lakdawalla DN, Phelps CE. Health technology assessment with diminishing returns to health: the generalized risk-adjusted cost-effectiveness (GRACE) approach. Value Health. 2021;24(2):244–9.
34. Lakdawalla DN, Doshi JA, Garrison LP Jr, Phelps CE, Basu A, Danzon PM. Defining elements of value in health care: a health economics approach: an ISPOR Special Task Force Report [3]. Value Health. 2018;21(2):131–9.
35. Rubin P, Arndorfer S, Lakdawalla D, Incerti D, Phelps CE, Masia N. Getting the math right when measuring the value of new medicines. EntityRisk. 2023. https://nopatientleftbehind.docsend.com/view/889u6zs74tra9x4a.
36. Shafrin J, Kim J, Cohen JT, Garrison LP, Goldman DA, Doshi JA, et al. Valuing the societal impact of medicines and other health technologies: a user guide to current best practices. Forum Health Econ Policy. 2024;27(1):29–116.
37. Panchal R, Nguyen D, Ghule P, Li N, Giannouchos T, Pan RJ, et al. Understanding patient cost-sharing thresholds for diabetes treatment attributes via a discrete choice experiment. J Manag Care Spec Pharm. 2023;29(2):139–50.
38. Pauly MV. The questionable economic case for value-based drug pricing in market health systems. Value Health. 2017;20(2):278–82.
39. Brownlee S, Wennberg J, Barry M, Fisher ES, Bynum J, Goodman D. Improving patient decision-making in health care: a 2012 Dartmouth Atlas report highlighting the East South Central Region. 2012. https://data.dartmouthatlas.org/downloads/reports/05_Decision_Making_East_South_Central_Region.pdf. Accessed 16 Dec 2024.
40. Silber JH, Rosenbaum PR, Niknam BA, Ross RN, Reiter JG, Hill AS, et al. Comparing outcomes and costs of medical patients treated at major teaching and non-teaching hospitals: a national matched analysis. J Gen Intern Med. 2020;35(3):743–52.
41. Fronsdal TL, Bhattacharya J, Tamang S. Variation in health care prices across public and private payers. National Bureau of Economic Research Working Paper Series. 2020; No. 27490.
42. Chernew ME, Hicks AL, Shah SA. Wide state-level variation in commercial health care prices suggests uneven impact of price regulation. Health Aff (Millwood). 2020;39(5):791–9.

43. Huang ES, Shook M, Jin L, Chin MH, Meltzer DO. The impact of patient preferences on the cost-effectiveness of intensive glucose control in older patients with new-onset diabetes. Diabetes Care. 2006;29(2):259–64.
44. Mattingly TJ 2nd, Hyman DA, Bai G. Pharmacy benefit managers: history, business practices, economics, and policy. JAMA Health Forum. 2023;4(11): e233804.
45. National Institute for Health and Care Excellence (NICE). Draft guidance consultation: tirzepatide for managing overweight and obesity (TA11156). 2024. https://www.nice.org.uk/guidance/GID-TA11156/documents/draft-guidance-2. Accessed 13 Aug 2024.
46. Gómez Lumbreras A, Tan MS, Villa-Zapata L, Ilham S, Earl JC, Malone DC. Cost-effectiveness analysis of five anti-obesity medications from a US payer's perspective. Nutr Metab Cardiovasc Dis. 2023;33(6):1268–76.
47. Dixit NM, Parikh NU, Ziaeian B, Fonarow GC. Economic modeling analysis of an intensive GDMT optimization program in hospitalized heart failure patients. Circ Heart Fail. 2023;16(12): e011218.
48. Grutters JP, Sculpher M, Briggs AH, Severens JL, Candel MJ, Stahl JE, et al. Acknowledging patient heterogeneity in economic evaluation: a systematic literature review. Pharmacoeconomics. 2013;31(2):111–23.
49. Willke RJ, Zheng Z, Subedi P, Althin R, Mullins CD. From concepts, theory, and evidence of heterogeneity of treatment effects to methodological approaches: a primer. BMC Med Res Methodol. 2012;13(12):185.
50. Shields GE, Clarkson P, Bullement A, Stevens W, Wilberforce M, Farragher T, et al. Advances in addressing patient heterogeneity in economic evaluation: a review of the methods literature. Pharmacoeconomics. 2024;42(7):737–49.
51. Ridley M, Schultz E, Chapman RH, Lai T. Fulfilling the promise of equity in value-based care: a focus on data and methods in health technology assessment: innovation and value initiative. Alexandria (VA); 2024. https://thevalueinitiative.org/wp-content/uploads/2024/04/2024-HEI-Data-and-Methods-Report_FINAL_Accessible.pdf.
52. Mertens D. Mixed methods design in evaluation. Thousand Oaks: SAGE Publications, Inc.; 2017.
53. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). REMS logic model: a framework to link program design with assessment. Guidance for Industry 2024.
54. Cartun Z, Kunz WG, Heinzerling L, Tomsitz D, Guertler A, Westphalen CB, et al. Cost-effectiveness of response-adapted de-escalation of immunotherapy in advanced melanoma. JAMA Dermatol. 2022;158(12):1387–93.
55. 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(5):415–30.
56. Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. Health Econ. 2004;13(5):461–75.
57. Nixon RM, Thompson SG. Methods for incorporating covariate adjustment, subgroup analysis and between-centre differences into cost-effectiveness evaluations. Health Econ. 2005;14(12):1217–29.
58. Bonander C, Svensson M. Using causal forests to assess heterogeneity in cost-effectiveness analysis. Health Econ. 2021;30(8):1818–32.
59. Moler-Zapata S, Grieve R, Lugo-Palacios D, Hutchings A, Silverwood R, Keele L, et al. Local instrumental variable methods to address confounding and heterogeneity when using electronic health records: an application to emergency surgery. Med Decis Making. 2022;42(8):1010–26.
60. Cao Q, Buskens E, Hillege HL, Jaarsma T, Postma M, Postmus D. Stratified treatment recommendation or one-size-fits-all? A health economic insight based on graphical exploration. Eur J Health Econ. 2019;20(3):475–82.
61. Goto D, Shih YT, Lecomte P, Olson M, Udeze C, Park Y, et al. Regression-based approaches to patient-centered cost-effectiveness analysis. Pharmacoeconomics. 2017;35(7):685–95.
62. Fletcher C, Chuang-Stein C, Paget MA, Reid C, Hawkins N. Subgroup analyses in cost-effectiveness analyses to support health technology assessments. Pharm Stat. 2014;13(4):265–74.
63. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Econ. 2006;15(12):1295–310.
64. Zaric GS. The impact of ignoring population heterogeneity when Markov models are used in cost-effectiveness analysis. Med Decis Making. 2003;23(5):379–96.
65. Spiegelhalter DJ, Best NG. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. Stat Med. 2003;22(23):3687–709.
66. Heeg BM, Damen J, Buskens E, Caleo S, de Charro F, van Hout BA. Modelling approaches: the case of schizophrenia. Pharmacoeconomics. 2008;26(8):633–48.
67. Sculpher M, Gafni A. Recognizing diversity in public preferences: the use of preference sub-groups in cost-effectiveness analysis. Health Econ. 2001;10(4):317–24.
68. Briggs ACK, Sculpher M. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.
69. Murphy P, Claxton L, Hodgson R, Glynn D, Beresford L, Walton M, et al. Exploring heterogeneity in histology-independent technologies and the implications for cost-effectiveness. Med Decis Making. 2021;41(2):165–78.
70. Vemer P, Goossens LM, Rutten-van Mölken MP. Not simply more of the same: distinguishing between patient heterogeneity and parameter uncertainty. Med Decis Making. 2014;34(8):1048–58.
71. Coyle D, Buxton MJ, O'Brien BJ. Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. Health Econ. 2003;12(5):421–7.
72. Basu A, Meltzer D. Value of information on preference heterogeneity and individualized care. Med Decis Making. 2007;27(2):112–27.
73. Espinoza MA, Manca A, Claxton K, Sculpher MJ. The value of heterogeneity for cost-effectiveness subgroup analysis: conceptual framework and application. Med Decis Making. 2014;34(8):951–64.
74. Jiao B, Carlson JJ, Garrison LP Jr, Basu A. Evaluating policies of expanding versus restricting first-line treatment choices: a cost-effectiveness analysis framework. Value Health. 2024;27(4):433–40.
75. Basu A, Carlson JJ, Veenstra DL. A framework for prioritizing research investments in precision medicine. Med Decis Making. 2016;36(5):567–80.
76. Saramago P, Espinoza MA, Sutton AJ, Manca A, Claxton K. The value of further research: the added value of individual-participant level data. Appl Health Econ Health Policy. 2019;17(3):273–84.
77. Hagan LM, Sulkowski MS, Schinazi RF. Cost analysis of sofosbuvir/ribavirin versus sofosbuvir/simeprevir for genotype 1 hepatitis C virus in interferon-ineligible/intolerant individuals. Hepatology. 2014;60(1):37–45.
78. Liu S, Watcha D, Holodniy M, Goldhaber-Fiebert JD. Sofosbuvir-based treatment regimens for chronic, genotype 1 hepatitis C virus infection in U.S. incarcerated populations: a cost-effectiveness analysis. Ann Intern Med. 2014;161(8):546–53.

79. Kazi DS, Garber AM, Shah RU, Dudley RA, Mell MW, Rhee C, et al. Cost-effectiveness of genotype-guided and dual antiplatelet therapies in acute coronary syndrome. Ann Intern Med. 2014;160(4):221–32.
80. Krashin JW, Edelman AB, Nichols MD, Allen AJ, Caughey AB, Rodriguez MI. Prohibiting consent: what are the costs of denying permanent contraception concurrent with abortion care? Am J Obstet Gynecol. 2014;211(1):76.e1-10.
81. Ding D, Hu H, Li S, Zhu Y, Shi Y, Liao M, et al. Cost-effectiveness analysis of durvalumab plus chemotherapy in the first-line treatment of extensive-stage small cell lung cancer. J Natl Compr Canc Netw. 2021;19(10):1141–7.
82. Ruiz Vargas E, Sposato LA, Lee SAW, Hachinski V, Cipriano LE. Anticoagulation therapy for atrial fibrillation in patients with Alzheimer's disease. Stroke. 2018;49(12):2844–50.
83. Logan C, Yumul I, Cepeda J, Pretorius V, Adler E, Aslam S, et al. Cost-effectiveness of using hepatitis C viremic hearts for transplantation into HCV-negative recipients. Am J Transplant. 2021;21(2):657–68.
84. Pandya A, Gupta A, Kamel H, Navi BB, Sanelli PC, Schackman BR. Carotid artery stenosis: cost effectiveness of assessment of cerebrovascular reserve to guide treatment of asymptomatic patients. Radiology. 2015;274(2):455–63.
85. Clemens A, Peng S, Brand S, Brueckmann M, Kansal A, Lim J, et al. Efficacy and cost-effectiveness of dabigatran etexilate versus warfarin in atrial fibrillation in different age subgroups. Am J Cardiol. 2014;114(6):849–55.
86. ISPOR. ISPOR science strategy real-world evidence. https://www.ispor.org/docs/default-source/ispor-science-strategy/ispor-science-strategy_online_00102242021.pdf?sfvrsn=e321a85_6. Accessed 25 Jul 2024.
87. Ramaekers BL, Joore MA, Grutters JP. How should we deal with patient heterogeneity in economic evaluation: a systematic review of national pharmacoeconomic guidelines. Value Health. 2013;16(5):855–62.
88. Dolan P. Effect of age on health state valuations. J Health Serv Res Policy. 2000;5(1):17–21.

Authors and Affiliations

Michael S. Willis[1] · Andreas Nilsson[1] · Cheryl A. Neslusan[2]

✉ Michael S. Willis
michael.willis@ihe.se

[1] The Swedish Institute for Health Economics (IHE), Lund, Sweden

[2] Janssen Scientific Affairs, LLC, a Johnson & Johnson Company, Titusville, New Jersey, USA