One Size Does Not Always Fit All in Value Assessment

Anirban Basu, PhD; Richard Grieve, PhD; Daryl Pritchard, PhD; and Warren Stevens, PhD

Rising healthcare costs have led to the emergence of a host of value frameworks aimed at both defining and quantifying what value means for healthcare in the United States.[1] Healthcare organizations, patient advocacy groups, and think tanks across the country have developed such frameworks to assess the potential value of new therapies.[2,3] In the United States, the framework to address new drug evaluation and pricing developed by the Institute for Clinical and Economic Review (ICER) has caught the attention of private payers.[4] Most recently, CVS Caremark announced plans to use the results from ICER's cost-effectiveness assessments to guide formulary decision making, which could lead to the exclusion of some high-cost drugs from some of its plans.[5]

The ICER perspective on what value means in healthcare—and some of the core methodology that it uses to evaluate alternative technologies—is based on long-standing academic concepts about cost-effectiveness analyses. These have been used in decision making outside the United States, notably by the likes of the National Institute for Health and Care Excellence in England and Wales and the Pharmaceutical Benefits Advisory Committee in Australia. However, these approaches have faced criticism,[6] not least because of the lack of attention given to heterogeneity in relative effectiveness and cost-effectiveness according to patients' characteristics and preferences.[7]

The Second Panel on Cost-Effectiveness in Health and Medicine called for heterogeneity to be considered through the presentation of subgroup-specific cost-effectiveness, where appropriate evidence exists.[8] Yet comparative and cost-effectiveness analyses have been slow to recognize heterogeneity and tend not to present subgroup value estimates.[9] By focusing on evaluating the overall average effectiveness, these value frameworks do not encourage the generation of useful evidence on heterogeneity that can inform differential decisions about the extent to which particular subgroups may benefit from new, high-cost healthcare technologies.

In most published value assessments, globally, heterogeneity has not been featured strongly in the reports of the main clinical results, and in the cost-effectiveness analysis these are addressed post hoc, after the main model has been built. For example, ICER's Evidence

Rating Matrix[10] makes no mention of whether a study attempts to detect or understand heterogeneity or report results by subgroup. There are genuine reasons to ignore heterogeneity in the absence of evidence, while there are cases where heterogeneity is ignored even with reliable evidence. ICER reports highlight both such cases.

One example in which evidence on heterogeneity could have been incorporated was ICER's report on treatments for rheumatoid arthritis (RA). ICER stated, "RA remains a remarkably complex disease to diagnose and manage. There are multiple phenotypic and genotypic variations in the pathogenesis of the disease that affect both the course of RA and the outcome of therapy."[11] Still, no attempt was made to evaluate cost-effectiveness of different therapeutic agents for subgroups. It is important to note that there was no direct evidence about treatment-effect heterogeneity across subgroups in any of the trials that were identified for the report. However, evidence beyond those trials clearly suggested that for patients receiving the control regimen, clinical responses differed according to age and functional status.[12] Hence, even if the relative effect of a new targeted immune modulator was constant across subgroups, there could still be substantial variation in the absolute effect scale required for estimates of cost-effectiveness.[13] It is not clear how incorporating such heterogeneity might have changed the overall assessment, but at the least, it could have triggered a different conversation around value for certain groups of patients. More generally, ignoring heterogeneity could result in therapies that may be highly effective and cost-effective for one particular group of patients not receiving coverage and reimbursement because they are not cost-effective for everyone.

In contrast, when evaluating programmed cell death 1 receptor agents in the treatment of non–small cell lung cancer, ICER's analyses relied on phase 2 and 3 trials that often did not have the power to establish subgroup effects reliably.[14] Therefore, despite emerging practice-based evidence that testing the level of programmed cell death ligand 1 protein that a tumor expresses can significantly help determine which patients may benefit from treatment, there was no reliable evidence during the clinical trial stage of development to model treatment-effect heterogeneity and report subgroup analyses.

These cases demonstrate 2 key barriers to driving greater reflection of heterogeneity in policy choices: positioning and availability of sources. The first case shows that even when clear evidence of heterogeneity of effect is present in published evidence, it is not moved to the front of the conversation, perhaps because it was not directly studied in the regulatory trial contexts. The second example points to the limitations when there is a distinct lack of strong empirical evidence on heterogeneity at the time clinical trials are conducted, but such evidence does emerge in clinical practice.

Implications of the Failure to Account for Heterogeneity in Value Assessment

It is imperative that value assessments encourage the recognition of evidence on heterogeneity for 2 reasons. First, it is well established that generating and reporting differential value assessment across subgroups leads to substantial health gains, both through treatment selection and coverage.[15-17] This means that simply providing value assessments for overall populations—even when clinical evidence shows differential effectiveness across subpopulations—leads to a disconnect between the assessment of evidence by payers versus clinicians and patients. This disconnect can ultimately lead to inefficient decision making around reimbursement and pricing.

Second, the recognition of clinical effect heterogeneity in value assessments can incentivize the production of better evidence on heterogeneity in the future. Greater availability of such evidence is critical to optimizing the benefits of a technology in the population. The current paucity of evidence around heterogeneity in the effectiveness of new treatments could reflect the lack of incentives to generate this information for regulatory purposes. By honoring and highlighting evidence on heterogeneity, value assessments could change that narrative and promote the generation of evidence.

Strategies for Accounting for Heterogeneity

Subgroup-specific value assessment may be an efficient strategy for accounting for heterogeneity on costs and benefits. Some value assessment framework developers have been reluctant to consider analysis of cost-effectiveness by subgroup or individual, largely for either practical reasons (eg, the concern that such granularity in results will require far larger studies) or statistical reasons (eg, potential selection bias, concerns about the greater risk of false positives or the relaxing of the standards of certainty). However, there are a growing number of approaches geared to overcoming these concerns that have been investigated with some success in recent years.

For example, the use of instrumental variables to define individual-level treatment effects of various approaches to treat prostate cancer used the same data sets that were used to produce the populationwide estimates of effectiveness that determined policy in the United States.[18] There have also been studies looking at various approaches to the use of propensity score matching to minimize selection bias in estimating cost-effectiveness by subgroups in the treatment of sepsis in the United Kingdom.[19] Bayesian modeling approaches, in which multiple sources of data on the relationship between the characteristics of patients and their risks are considered, have also been suggested for more effective interpretation of potential subgroup effects.[19,20]

These approaches have all shown how value assessment can be conducted in a way that better accounts for heterogeneity of treatment effect, and they highlight the need for a more nuanced view of the evidence hierarchy—one that recognizes a greater role for real-world data as a complement to traditional randomized controlled trial designs. The practice of cost-effectiveness analysis is mostly focused on the simple comparison of population-based treatment impacts. However, decision makers may benefit from knowing how these impacts vary across subsets of the population so that benefit designs or coverage decisions are aligned in an increasingly complex healthcare delivery system that is rapidly evolving toward an increased use of personalized medicine. One step forward in this evolution may occur if value assessments open a conversation about evidence of heterogeneity of effect with manufacturers during their initial interactions around sharing public and proprietary evidence on new drugs. No one is recommending that a health technology assessment body make statements based on nonexistent evidence, but where it exists, it should not be ignored or made a footnote; patients deserve better.

Conclusions

Laying a clear path for incorporating reliable evidence on heterogeneity in value assessments could improve its applicability for healthcare decision making. This could include not reporting population average cost-effectiveness results when there are distinct differences in subgroup-specific results and sufficient accounting for such heterogeneity among patients. Importantly, creating an environment that respects and rewards evidence on heterogeneity should help value frameworks evolve to become more applicable and appropriate for payers' decisions and promote generation of evidence on heterogeneity.

The worlds of comparative effectiveness and cost-effectiveness research must catch up with the evolution of how healthcare is shifting its emphasis from addressing disease in populations to addressing disease in patients, both now and in the future. There is no better time to begin realizing this change in perspective than now. ∎

Author Affiliations: The CHOICE Institute, Departments of Pharmacy, Health Services, and Economics, University of Washington (AB), Seattle, WA; London School of Hygiene & Tropical Medicine (RG), London, United Kingdom; Personalized Medicine Coalition (DP), Washington, DC; Parexel International (WS), Durham, NC.

Address Correspondence to: Warren Stevens, PhD, Parexel International, 2520 Meridian Pkwy, Ste 200, Durham, NC 27713. Email: wstevens15@gmail.com.

COMMENTARY

REFERENCES

1. Kesselheim AS, Avorn J, Sarpatwari A. The high cost of prescription drugs in the United States: origins and prospects for reform. *JAMA*. 2016;316(8):858-871. doi: 10.1001/jama.2016.11237.

2. Wilson L, Lin T, Wang L, et al. Evaluation of the ASCO value framework for anticancer drugs at an academic medical center. *J Manag Care Spec Pharm*. 2017;23(2):163-169. doi: 10.18553/jmcp.2017.23.2.163.

3. Schnipper LE, Davidson NE, Wollins DS, et al; American Society of Clinical Oncology. American Society of Clinical Oncology statement: a conceptual framework to assess the value of cancer treatment options. *J Clin Oncol*. 2015;33(23):2563-2577. doi: 10.1200/JCO.2015.61.6706.

4. ICER launches new drug assessment program with $5.2M from Arnold Foundation [news release]. Houston, TX, and Boston, MA: Arnold Foundation; July 21, 2015. arnoldventures.org/newsroom/icer-launches-new-drug-assessment-program-with-5-2-million-award-from-the-laura-and-john-arnold-foundation. Accessed September 16, 2019.

5. Current and new approaches to making drugs more affordable. CVS Health website. cvshealth.com/sites/default/files/cvs-health-current-and-new-approaches-to-making-drugs-more-affordable.pdf. Published August 2018. Accessed September 16, 2019.

6. Birch S, Gafni A. On being NICE in the UK: guidelines for technology appraisal for the NHS in England and Wales. *Health Econ*. 2002;11(3):185-191.

7. 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):951-964. doi: 10.1177/0272989X14538705.

8. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second Panel on Cost-Effectiveness in Health and Medicine [erratum in *JAMA*. 2016;316(18):1924. doi: 10.1001/jama.2016.15518]. *JAMA*. 2016;316(10):1093-1103. doi: 10.1001/jama.2016.12195.

9. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages [erratum in *Milbank Q*. 2006;84(4):759-760]. *Milbank Q*. 2004;82(4):661-687. doi: 10.1111/j.0887-378X.2004.00327.x.

10. Ollendorf D, Pearson SD. ICER evidence rating matrix: a user's guide. Institute for Clinical and Economic Review website. icer-review.org/wp-content/uploads/2013/04/Rating-Matrix-User-Guide-Exec-Summ-FINAL.pdf. Published 2013. Accessed September 16, 2019.

11. Targeted immune modulators for rheumatoid arthritis: effectiveness & value. Institute for Clinical and Economic Review website. icer-review.org/wp-content/uploads/2018/08/NE_CEPAC_RA_Evidence_Report_FINAL_040717.pdf. Published April 7, 2017. Accessed September 16, 2019.

12. Hetland ML, Christensen IJ, Tarp U, et al; All Departments of Rheumatology in Denmark. Direct comparison of treatment responses, remission rates, and drug adherence in patients with rheumatoid arthritis treated with adalimumab, etanercept, or infliximab: results from eight years of surveillance of clinical practice in the nationwide Danish DANBIO registry. *Arthritis Rheum*. 2010;62(1):22-32. doi: 10.1002/art.27227.

13. Kent DM, Nelson J, Dahabreh IJ, Rothwell PM, Altman DG, Hayward RA. Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. *Int J Epidemiol*. 2016;45(6):2075-2088. doi: 10.1093/ije/dyw118.

14. Treatment options for advanced non-small cell lung cancer: effectiveness, value and value-based price benchmarks. Institute for Clinical and Economic Review website. icer-review.org/wp-content/uploads/2016/08/MWCEPAC_NSCLC_Evidence_Report_Plus_Supplement_101716.pdf. Updated October 7, 2016. Accessed September 16, 2019.

15. Basu A. Economics of individualization in comparative effectiveness research and a basis for a patient-centered health care. *J Health Econ*. 2011;30(3):549-559. doi: 10.1016/j.jhealeco.2011.03.004.

16. Basu A, Jena AB, Philipson TJ. The impact of comparative effectiveness research on health and health care spending. *J Health Econ*. 2011;30(4):695-706. doi: 10.1016/j.jhealeco.2011.05.012.

17. 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):951-964. doi: 10.1177/0272989X14538705.

18. Basu A. Estimating person-centered treatment (PeT) effects using instrumental variables: an application to evaluating prostate cancer treatments. *J Appl Econ (Chichester Engl)*. 2014;29(4):671-691. doi: 10.1002/jae.2343.

19. Kreif N, Grieve R, Radice R, Sadique Z, Ramsahai R, Sekhon JS. Methods for estimating subgroup effects in cost-effectiveness analyses that use observational data. *Med Decis Making*. 2012;32(6):750-763. doi: 10.1177/0272989X12448929.

20. Heckman JJ, Lopes HF, Piatek R. Treatment effects: a Bayesian perspective. *Econom Rev*. 2014;33(1-4). doi: 10.1080/07474938.2013.807103.

Value Assessment and Heterogeneity: Another Side to the Story

Steven D. Pearson, MD, MSc
Institute for Clinical and Economic Review
Boston, MA

To assert that value assessment is at fault for ignoring heterogeneity in relative effectiveness, or for minimizing the importance of subgroups (the relationship between the heterogeneity and subgroups being critical but often obscured), is a bit like finding a man building a house out of the wood he can find or borrow from neighbors and criticizing him for not using bricks that no one will sell him. He has a need for shelter; he does the best he can with the resources he can get; and he would love to have bricks, but powers beyond his control make that impossible.

Let's start with the goal of value assessment. What are we trying to build? Is the aim to provide a tool to help inform the clinical care of individual patients? Value assessment can indeed be oriented to serve the interests of enhanced shared decision making for individual patients.[1] Evidence-based tools can help frame the many different elements of clinical decisions that are important to patients and provide summaries of evidence from population averages or, ideally, from results for patients with similar clinical characteristics. What most distinguishes this form of patient-targeted value assessment is that it keeps all the various elements of risks, benefits, and other elements of value disaggregated so patients can place their own unique "weights" on them and add them up or otherwise consider them in some quasi or formally quantitative process.

This is value assessment in service of what I would call individual heterogeneity—the variation among individual patients that a skilled clinician can illuminate and apply to tailor the care for a patient in their best interest. Individual patients will have unique clinical, emotional, social, and

other characteristics that providers should always consider to help select the "best" drug or other treatment option, a critical goal of good medical care.

But there is a second kind of heterogeneity that can be called population heterogeneity, and this is the home territory of the value assessment performed by health technology assessment (HTA) agencies and research groups around the world. The goal of HTA here is not to inform individual clinical decisions but to inform the decisions that are taken at the population level: coverage and pricing. The heterogeneity that matters most in these decisions reflects variation in outcomes at a higher level than the individual and has two forms: one that is knowable in advance of treatment and one that is unknowable to the patient and clinician because its causes are unknown to all. In the latter case, evidence may show that patient outcomes appear something like a bell curve, with some patients receiving "average" benefits and harms, whereas others experience better or worse outcomes. The key feature of unknowable population heterogeneity is that there are no signposts, biomarkers, or key clinical indicators that can helpfully predict whether a specific patient will have average outcomes. Although it is still helpful to understand the range of outcomes for different patients, unknowable population heterogeneity leaves patients, clinicians, and policy makers largely reliant on population averages.

However, sometimes evidence can provide a guide to help identify when patients can be expected to experience relatively better or worse outcomes. And here lies the connection to subgroup analysis. Formal subgroup analysis is the most powerful way for HTA to identify how the risks and benefits of treatment may vary systematically within a larger population.[2] I would argue that it is misleading to claim that HTA has been slow or recalcitrant in recognizing, seeking, and applying subgroup information to create precise value assessments at the population level. Seeking subgroups for which a drug might be most effective and cost-effective is a vigorous part of HTA. As one example, at the National Institute for Health and Care Excellence in the United Kingdom, this effort leads the agency to designate positive funding decisions for subgroups for many drugs that would otherwise fail a general test of cost-effectiveness across the entire labeled population.[3]

For my HTA agency, the Institute for Clinical and Economic Review (ICER), the hunt for subgroups has led us to create stratified cost-effectiveness findings for different patient subgroups in reviews for treatments such as proprotein convertase subtilisin/kexin type 9 drugs for hypercholesterolemia, Spinraza and Zolgensma treatments for spinal muscular atrophy, and preventive treatments for migraine.[4] In many cases, we

have found strikingly different cost-effectiveness data across subgroups. We are eager for more data and omnivorous in our appetite for evidence from various sources. In the US healthcare system, where there is a single price for a drug regardless of its use, we pursue the use of subgroups to the point of calculating separate value-based price benchmarks for each subpopulation, even though we must also recognize the reality of the US system and do a weighted calculation across all subgroups to calculate a single population value-based price.

Why have we not been able to identify and model subpopulations within every review we have done? The simple answer is that data are often not available on both the treatment of interest and its main comparator that can be used to assess subgroup effects, but there is a deeper and more complex issue in play. Drug makers face very conflicting incentives in helping to identify subpopulations who might have more—or less—benefit from their drug. On one hand, a special niche within a broader label in which patients can be shown to have superior outcomes might help the drug maker compete for market share against other drugs. Conversely, slicing the data from the overall labeled population into subpopulations might show diminished benefit in large swaths of the patient population, helping clinicians and payers limit use of the drug to the narrower subgroups that benefit most within the broader label. For that reason, and perhaps others, even though we at ICER continue to make routine requests to drug makers at the initiation of every review for stratified or patient-level data, we continue to routinely have this request unfulfilled. Whether we, like the man building his house who would love to be able to do so with bricks, should be criticized for lacking bricks and doing the best we can with the wood we can cobble together, is dubious.

Ultimately, when value assessment seeks to build an analysis to inform population-level policies, it will never be able to consider the granular details that make each patient unique. But the middle ground of subgroup analysis is ripe for common efforts and an area in which we can hope to see progress in our ability to generate, analyze, and apply data in service of more sophisticated coverage and pricing policies. ∎

REFERENCES

1. Patient perspective value framework. Milken Institute FasterCures website. fastercures.org/programs/patients-count/patient-perspective-value-framework. Accessed October 14, 2019.
2. 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):951-964. doi: 10.1177/0272989X14538705.
3. Barham L. NICE numbers at 20. PharmaPhorum website. pharmaphorum.com/views-analysis-market-access/nice-numbers-20. Published July 18, 2019. Accessed October 14, 2019.
4. Topics. Institute for Clinical and Economic Review website. icer-review.org/topics. Accessed October 14, 2019.