Published in final edited form as:

Med Decis Making. 2018 May ; 38(4): 487–494. doi:10.1177/0272989X17746989.

Patient variability seldom assessed in cost-effectiveness studies

Tara A. Lavelle, PhD¹, David Kent, MD, MSc², Christine Lundquist, MPH², Teja Thorat, MPH¹, Joshua T. Cohen, PhD¹, John B. Wong, MD³, Natalia Olchanski, MS¹, Peter J. Neumann, ScD¹

¹Center for the Evaluation of Value and Risk in Health, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA

²Predictive Analytics and Comparative Effectiveness (PACE) Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA

³Division of Clinical Decision Making, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA

Abstract

Background—Cost-effectiveness analysis (CEA) estimates can vary substantially across patient subgroups when patient characteristics influence preferences, outcome risks, treatment effectiveness, life expectancy or associated costs. However, no systematic review has reported on the frequency of subgroup analysis in CEA, what type of heterogeneity they address, and how often heterogeneity influences whether cost-effectiveness ratios exceed or fall below conventional thresholds.

Methods—We reviewed the CEA literature catalogued in the Tufts Medical Center CEA Registry, a repository describing cost-utility analyses published through 2015. After randomly selecting 200 of 642 articles published in 2014, we ascertained whether each study reported subgroup results and collected data on the defining characteristics of these subgroups. We identified whether any of the CEA subgroup results crossed conventional cost-effectiveness benchmarks (e.g., \$100,000 per QALY), and compared characteristics of studies with and without subgroup-specific findings.

Results—Thirty-eight studies (19%) reported patient subgroup results. Articles reporting subgroup analyses were more likely to be U.S.-based, government funded (vs. drug industry- or nonprofit foundation-funded), and to focus on primary or secondary (vs. tertiary) prevention ($p < 0.05$ for comparisons). One or more patient characteristic was used to stratify CEA results 68 times within the 38 studies, with most stratifications using one characteristic ($n = 47$), most commonly age ($n = 35$). Among the 23 stratifications reported alongside average ratios in U.S. studies, 13 produced subgroup ratios that crossed a conventional CEA ratio benchmark.

Conclusions—Most CEAs do not report any subgroup results and those that do most often stratify only by patient age. Over one-half of subgroup analyses reported could lead to different value based decision-making for at least some patients.

Keywords

Cost-effectiveness; heterogeneity; subgroups

Background

Cost-effectiveness analyses (CEAs) characterize the value of health care programs by comparing the costs of alternative interventions with the respective health outcomes realized, and can be used to help decision makers prioritize the best use of health care dollars. CEAs help to inform national priority setting for health care in many countries, including the UK, Canada, Australia, France, and Norway. While U.S. government health care programs, with some exceptions, generally do not use CEA information for coverage decisions, use of this information does exert some influence on public health and clinical practice guidelines the U.S.¹

The cost-effectiveness of an intervention may depend on the characteristics of the individuals to whom it is targeted, i.e. whether the patients have co-morbid conditions, or other observable characteristics that may influence their life expectancy, treatment or outcome preferences, individual outcome risks, intervention benefits and harms, or individual health care costs. However, many CEAs report incremental costs and benefits of an intervention across a range of individuals, on average, and without regard to specific and observable characteristics of individual patients.

Better characterizing how cost-effectiveness varies by observable patient characteristics could help to identify which patients should be targeted at the time of a treatment decision based on their characteristics. Such an approach could help mitigate wasteful or inefficient spending on patients for whom a treatment confers minimal value, even for treatments that are worthwhile on average.^{2–4} Conversely, physicians could recommend interventions that confer high value for a subset of patients, even when evidence shows that the intervention is not worthwhile on average. Although studies now more than a decade old have highlighted the need for more stratified and personalized cost-effectiveness information,⁵ no review has systematically elucidated the frequency with which this recommendation has been incorporated into CEAs, and the implications for differential decision-making.

This paper aimed to review the CEA literature to determine: 1) how often CEAs estimate cost-effectiveness variation by identifiable patient characteristics; 2) the characteristics of studies (i.e., country of study, disease area, funding source) that incorporate patient heterogeneity; and 3) how often patient subgroup results would lead to targeted recommendations that differ from recommendations based on average, according to conventional CE thresholds reflecting “reasonable” value for money.

Methods

Data

The Tufts Medical Center CEA Registry (www.cearegistry.org)⁶ catalogs selected information from all English-language cost utility analyses (CUAs) published through

2015.⁷ CUAs are CEAs that quantify health benefits in terms of quality adjusted life years (QALYs). We describe the CEA Registry search strategy and data collection methods elsewhere.⁶ Briefly, Registry reviewers searched Medline for English-language articles using the keywords “QALY”, “quality-adjusted” and “cost-utility analysis.” Researchers screened retrieved abstracts to identify studies reporting original cost-effectiveness estimates expressed using QALYs. The search excluded reviews, editorials and other methodological studies. Two independent reviewers screen selected papers in full to extract information on study methods, incremental cost-effectiveness ratios (ICERs) and health state preference (i.e., “utility”) weights. The CEA Registry catalogs more than 5,600 English language CUAs and more than 14,400 cost-effectiveness ratios.

Review and Analysis

We used only the CUAs contained in the Registry for this study and thus relied on previously extracted study methods and results information. Due to the time it took to review articles and to abstract additional data not already contained in the Registry, we randomly selected 200 CUAs from the 642 articles catalogued in the Tufts CEA Registry from 2014, the most recent year of data in the Registry when this study began. For each study in our review, we extracted data contained in the CEA Registry on the type of intervention being evaluated, disease classification, country of study, funding source, and prevention stage. Using pre-specified data collection forms, two trained readers reviewed each article to determine whether it reported cost-effectiveness results for subgroups characterized by one or more identifiable patient characteristic. Among those studies that report results by patient subgroups, we collected additional data from each article, including quality metrics based on previously published theoretical and methodological guidance on incorporating patient heterogeneity into CEAs.^{2,8,9} This information included whether the methods section of the paper specified the heterogeneity subgroup, whether each pre-specified patient subgroup had CEA results presented, the terms were used by authors to identify heterogeneity (i.e., subgroup, stratification or sensitivity analysis), whether the reported costs, QALYs and ICERs included estimates of statistical uncertainty, whether the authors discussed different interpretations of the value of the intervention for different patient subgroups, and whether the authors discussed any concerns about distributional equity related to heterogeneity.

We collected information on the observable clinical and socio-demographic characteristics that defined the patient subgroups (age, gender, race, severity of disease, presence of disease/comorbid conditions, lifestyle factors), and whether the study used these variables alone (univariate) or in combination (multivariate) to define subgroups. We also identified studies that stratified results based on patient characteristic predicted risk variables. These risk predictions included both previously developed scores, such as the Framingham risk score,¹⁰ and unique risk algorithms developed based on observable or measurable patient characteristics, in papers we reviewed.

Differences in identifiable clinical and socio-demographic patient characteristics can lead to variability in at least five key CEA inputs (Figure 1): 1) life expectancy, 2) baseline risk of health outcomes, 3) treatment effect, 4) patient preferences for treatment or outcomes; and 5) costs, including intervention costs, costs of future clinical events, and productivity costs.²

Patient subgroups within CEAs conducted alongside randomized controlled trials (RCTs) inherently capture differences across all domains, but may not measure or report all of these differences. Subgroups created within model-based CEAs would ideally be created to account for differences in model inputs for each subgroup along all domains, but the feasibility of such an approach would be dependent on the exactness of the model. We examined what potential sources of heterogeneity were accounted for in the subgroup analyses in our review, based on what was reported in the paper.

In the subset of reviewed studies that were conducted in the U.S., we compared subgroup ratios reported alongside average ratios to identify situations in which subgroup results would cross conventional CEA ratio U.S. benchmarks and could lead to different value-based decision making for certain patients. Specifically, we identified cases in which the stratification created a subgroup ratio that was *more* favorable than the average CEA result at a variety of thresholds: cost-saving (at least one subgroup is cost-saving, average is not); \$50,000/QALY (at least one subgroup <\$50,000/QALY, average is >\$50,000/QALY); \$100,000/QALY (at least one subgroup <\$100,000/QALY, average is >\$100,000/QALY); and dominated (at least one subgroup is not dominated even if the population average is). We also identified cases in which the stratification created a subgroup ICER that was *less* favorable than the average CEA result at the same thresholds (e.g., cost-saving (average is cost-saving, at least one subgroup is not), \$50,000/QALY (average <\$50,000/QALY, at least one subgroup is >\$50,000/QALY) and so forth. We adjusted all ICERs to 2015 USD dollars using the Bureau of Labor Statistics 2015 Consumer Price Index Report.¹¹

Following data collection, readers met with a third reviewer to resolve any discrepancies in data coding. We summarized review data with descriptive statistics. Using chi-squared tests to determine statistical significance, we compared Registry data from studies that did and did not include subgroups.

Results

According to our categorization, 38 of the 200 reviewed studies (19%) reported subgroup results based on observable or measurable patient information. Articles that incorporated patient heterogeneity were more likely to be U.S.-based studies, to include a government funder, and to focus on primary or secondary (vs. tertiary) prevention ($p<0.05$ for all comparisons; Table 1).

We found that one or more patient characteristic was used to stratify results 68 times within the 38 studies with subgroups (Table 2). The most common stratifying variable was patient age (35 stratifications), followed by a clinical (lab, test or procedure) result ($n=7$), and predicted risk ($n=6$). The vast majority of stratifications examined only one patient characteristic ($n=47$), while 15 stratifications used two or more characteristics to define subgroups, and 6 used a predicted risk score.

Although we expected patient subgroups to differ on at least five sources of variability (life expectancy, baseline risk, treatment effect, preferences and cost, as outlined above), few studies described how patient subgroups differed on more than one of these factors (Table

2), and no studies mentioned any sources of heterogeneity extending beyond these five pre-specified factors. For 35 stratifications, authors reported differentiating subgroups based on one source of heterogeneity, with differing baseline risks (n= 16) being the most common. Among the 68 stratifications we identified, only in 17 did the authors report more than one source of subgroup variability. For 16 stratifications, there was no mention of the sources of variability differentiating subgroups (Table 2).

When examining subgroup results from 23 stratifications that were reported alongside average ratios in 17 U.S. based studies, we found 13 cases in which the stratification produced at least one ratio that crossed conventional U.S. CEA ratio benchmarks as compared with the population average ratio. This means that for a particular intervention, over half the cases had a least some group of patients that should be receiving different treatment recommendations from the average patient, based on the value of the intervention. Specifically, the results from 8 stratifications created at least one subgroup ratio that was more favorable than the average CEA result, most frequently at the \$50,000/QALY threshold (Table 3). The results from 10 stratifications created at least one subgroup ratio that was less favorable than the average CEA result, most frequently at the \$100,000/QALY threshold. In 10 stratifications, subgroup ratios produced were all on the same side of conventional CEA ratio benchmarks as the average ratios (Table 3) and would not lead to different value-based decisions subgroup patients compared with the overall group.

The majority of papers with subgroups (n=32) pre-specified in the methods section that they would report results according to all of the patient characteristics ultimately used for stratification (Table 4), although the specific terms “subgroup” or “stratified” analysis was only mentioned in 9 of these papers. In the remaining papers, authors labeled the stratified analyses as “sensitivity” or “scenario” analyses (n=15) or did not use a term (n=14). All studies reported the CEA results separately for each pre-specified subgroup, but only 9 included some type of uncertainty estimate around subgroup results, either in the cost component of the CEA, the QALY, or both. Authors discussed different interpretations of the value of the intervention for different subgroups in 15 papers; no papers discussed equity concerns related to the reporting of subgroup results.

Discussion

Over the past decade, the CEA literature has grown rapidly,¹² thereby warranting greater attention to the methods used, including how patient heterogeneity is assessed and reported. Incorporating patient heterogeneity into CEAs and using targeted value information as part of the clinical decision making process might lead to substantial gains in efficiency, because interventions that are cost effective overall may be cost-ineffective in many patients to which they are applied. Our review indicates that most CEAs do not stratify results by observable patient characteristics, indicating that we currently do not have all the information we need to routinely target interventions to certain individuals and groups based on value. Among U.S. studies that did report heterogeneity, we found that over half of reported subgroup results would potentially lead to different value based decision making for at least some patients, compared to the average patient.

The lack of subgroup consideration in CEAs is problematic because population-level evidence is frequently insufficient to guide informed decision making at the individual level. For example, previous studies have shown that patients enrolled in RCTs often have large variation in baseline risk, which can have substantial impact on absolute treatment effects¹³ that could lead to very different cost-effectiveness and estimates or even harm exceeding benefit. In CEAs, the relevant patient variability that may influence CEA results extends beyond baseline risk to include life expectancy, relative treatment effect, patient preferences, and costs, all of which should define patient variability in subgroups. Among studies in our review that did include subgroups, most authors reported accounting for only one source of heterogeneity, most frequently baseline risk. While it is clear that this under-accounting of the sources of heterogeneity in analyses has the potential to bias the subgroup results,¹⁴ it is hard to infer the direction of this bias, because the different components of heterogeneity may operate in compensatory directions or synergistically.

The absence of patient subgroups in CEAs is not surprising, and likely reflects limitations in the primary literature and data upon which analyses are based. Limitations in the reporting of subgroups analysis in clinical trials has been repeatedly noted; there are serious concerns about the credibility of subgroup analysis, and many fundamental issues about the best way to conduct these analyses remain controversial and unsettled.^{15,16} Some clinical trials may be underpowered to detect differences in treatment effect within subgroups, particularly for rare diseases or for illnesses with small patient populations such as in children. Even when subgroups are conducted within RCTs, the results of these subgroup analyses are often not presented in sufficient detail to adequately create fully defined subgroup analyses in model-based CEAs.

Furthermore, attempts to balance the efficient use of resources with concerns for distributional equity have led to guidance in countries such as the UK that highlight the legal obligations of decision makers to avoid health resource prioritization that would discriminate against individuals on the basis of age, gender, disability, race/ethnicity, and other socio-demographic characteristics.¹⁷ We find that studies incorporating patient heterogeneity are more likely to originate from the U.S. This may reflect the less stringent regulatory concerns and restrictions regarding distributional equity in U.S. versus other countries. Notably, the Second Panel on Cost-effectiveness in Health and Medicine notes the ethical concerns in conducting CEAs targeted at certain groups of individuals, and recommends providing decision makers with sufficient information regarding equity issues alongside cost-effectiveness results,¹ which can be used to infer the appropriateness of targeting specific subgroups in clinical and reimbursement policies.

We found that the vast majority of CEA subgroups were created using only one observable patient variable. Previous research¹⁸ has highlighted several fundamental problems with the use of univariate subgroups in clinical research, and these concerns are relevant for cost-effectiveness research. Specifically, because each patient has innumerable characteristics, using univariate (one-variable-at-time) subgroups disaggregates results into a series of groups that are not mutually exclusive, such that one patient will fall into many different subgroups, each with potentially different results motivating potentially different decisions.¹⁹ One-variable-at-a-time subgroup analyses may therefore fail to capture the substantial

heterogeneity across individual patients, because they compare groups of patients that differ systematically only on a single characteristic (i.e. gender) while patients differ simultaneously across many characteristics (gender, race, history of disease, etc.). These limitations may be overcome in part by using multivariable subgroups and predicted risk models to define more patient-centric subgroups,²⁰ though these were used infrequently in the papers we reviewed.

Properly characterizing heterogeneity in CEAs can help decision-makers more efficiently target therapies and programs to certain individuals and groups, and not others because of low-value. Ongoing efforts to curb low-value care, such as the “Choosing Wisely” campaigns in the U.S. and Canada, have sought to limit the use of “unnecessary” health services by creating and disseminating lists of tests, treatments and procedures that generally provide little or no benefit to certain individuals, such as cardiac screening for low-risk, asymptomatic patients and pap smears for women under the age of 21.²¹ This work has been further refined and developed into claims based measures of low-value care, used by researchers to characterize and quantify the overuse of health care resources in the U.S.^{22–24} While important initiatives, these efforts generally define “low-value” care as care which provides little to no benefit, or has the potential for net harm in particular patient groups. The measures generally do not include information about services which should be avoided in subgroups of patients because the high costs of care outweigh the benefits. An expansion of work in this direction could lead to significant improvements in the efficiency of our health care system, but the results of our research show that such evidence generated from economic evaluations may not currently be sufficient to support such efforts. We did find that studies funded by a government source were more likely to include patient subgroups, compared to studies funded by private sources.

Guidance for researchers on when and how to incorporate heterogeneity into CEA has previously been published.^{2,8,9} These guidelines indicate that heterogeneity should be considered in the study design, analysis and reporting phases of the study. In the design phase, researchers should identify the observable patient characteristics to be used in subgroups, the sources of heterogeneity that will be incorporated, and justify planned subgroup analyses, noting equity considerations. In the analysis and reporting phases, analysts are urged to separately report point estimates, statistical precision for each subgroup result, discuss whether subgroup differences lead to different conclusions, cite implications for equity, and explore the feasibility of subgroup targeting for clinical and reimbursement policy. We used these guidelines to evaluate the quality of the papers included in our review, and noted that studies fell short on several domains. Specifically, published CEAs rarely reported the statistical precision of subgroup CEA results, rarely discussed the different interpretations of the results for subgroups vs. the average, and never discussed equity. Of note, we were not able to ascertain what information authors had pre-specified for their study, and instead evaluated what information was stated in their methods sections. Regardless, the shortcomings noted in the available data shed some light on the work needed to improve the quality of this important component of CEA.

Despite recommendations to report patient variability in CEAs, we find that few papers follow this advice, and those that do may lack the rigorous quality that would yield valid and

credible results. Specifically, most patient subgroup analyses report one-variable-at-time stratifications rather than examining the combined effects of multiple influential characteristics simultaneously as recommended.^{13,20} Furthermore, these analyses do not account for the multiple ways in which patients may differ from each other in the analysis. Nevertheless, among reported subgroup analyses, over half produce results that would lead to different value based decision-making for at least some patients. The full potential for subgroup analyses to inform value-based decisions remains largely unexplored, however. Improving the quality and utilization of subgroups in CEAs should be a focus of future research, as it may help decision-makers target health care resources more efficiently.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: Funding for this study was provided by the National Institutes of Health (NIH) Grant U01NS086294.

References

1. Neumann PJ, Ganiats TG, Russell LB, Sanders GD, Siegel JE. Cost-Effectiveness in health and Medicine. Second Edition: Oxford Scholarship Online; 2016.
2. Sculpher M Subgroups and heterogeneity in cost-effectiveness analysis. *Pharmacoeconomics*. 2008;26(9):799–806. [PubMed: 18767899]
3. Ioannidis JP, Garber AM. Individualized cost-effectiveness analysis. *PLoS Med*. 2011;8(7):e1001058. [PubMed: 21765810]
4. Kent DM, Vijan S, Hayward RA, Griffith JL, Beshansky JR, Selker HP. Tissue plasminogen activator was cost-effective compared to streptokinase in only selected patients with acute myocardial infarction. *J Clin Epidemiol*. 2004;57(8):843–852. [PubMed: 15485737]
5. Coyle D, Buxton MJ, O'Brien BJ. Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. *Health Econ*. 2003;12(5):421–427. [PubMed: 12720259]
6. Thorat T, Cangelosi M, Neumann PJ. Skills of the Trade: The Tufts Cost-Effectiveness Analysis Registry. *Journal of Benefit-Cost Analysis*. 2012;3(1):1–9.
7. Blog TMCCR. Cost-Effectiveness Analysis (CEA) Registry Data Now Updated Through 2015. 2016; <http://healtheconomics.tuftsmedicalcenter.org/cear4/Resources/CEARegistryBlog/tabid/69/EntryId/451/Cost-Effectiveness-Analysis-CEA-Registry-Data-Now-Updated-Through-2015.aspx>. Accessed March 16, 2017.
8. Grutters JP, Sculpher M, Briggs AH, et al. Acknowledging patient heterogeneity in economic evaluation : a systematic literature review. *Pharmacoeconomics*. 2013;31(2):111–123. [PubMed: 23329430]
9. 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):855–862. [PubMed: 23947981]
10. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97(18):1837–1847. [PubMed: 9603539]
11. Statistics. USDoL-BoL. Consumer Price Index: All Urban Consumers (CPI-U) Detailed Report. 2015; <https://www.bls.gov/cpi/cpid1512.pdf>. Accessed March 16, 2017.
12. Neumann PJ, Thorat T, Shi J, Saret CJ, Cohen JT. The changing face of the cost-utility literature, 1990–2012. *Value Health*. 2015;18(2):271–277. [PubMed: 25773562]
13. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA*. 2007;298(10):1209–1212. [PubMed: 17848656]

14. van Klaveren D, Vergouwe Y, Farooq V, Serruys PW, Steyerberg EW. Estimates of absolute treatment benefit for individual patients required careful modeling of statistical interactions. *J Clin Epidemiol*. 2015;68(11):1366–1374. [PubMed: 25814403]
15. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*. 2000;355(9209):1064–1069. [PubMed: 10744093]
16. Lagakos SW. The challenge of subgroup analyses--reporting without distorting. *N Engl J Med*. 2006;354(16):1667–1669. [PubMed: 16625007]
17. Culyer AJ, Bombard Y. An equity checklist: a framework for health technology assessments Centre for Health Economics, The University of York 2011.
18. Hayward RA, Kent DM, Vijan S, Hofer TP. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. *BMC Med Res Methodol*. 2006;6:18. [PubMed: 16613605]
19. Dahabreh IJ, Hayward R, Kent DM. Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. *Int J Epidemiol*. 2016.
20. Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials*. 2010;11:85. [PubMed: 20704705]
21. Foundation ABoIMA. Choosing Wisely: Promoting Conversations Between Providers and Patients. 2017; <http://www.choosingwisely.org/>. Accessed March 16, 2017.
22. Schwartz AL, Landon BE, Elshaug AG, Chernew ME, McWilliams JM. Measuring low-value care in Medicare. *JAMA Intern Med*. 2014;174(7):1067–1076. [PubMed: 24819824]
23. Chua KP, Schwartz AL, Volerman A, Conti RM, Huang ES. Use of Low-Value Pediatric Services Among the Commercially Insured. *Pediatrics*. 2016;138(6).
24. Schwartz AL, Zaslavsky AM, Landon BE, Chernew ME, McWilliams JM. Low-Value Service Use in Provider Organizations. *Health Serv Res*. 2016.

Figure 1.

How patient variability can be assessed in cost-effectiveness analyses

Table 1.

Study characteristics

	Heterogeneity Defined	
	Yes (n=38)	No (n=162)
<i>Diseases</i>		
Cancer	10 (26%)	34 (21%)
Infectious	8 (21%)	28 (17%)
Cardiovascular	10 (26%)	26 (16%)
Neuropsychiatric & Neurological	2 (5%)	19 (12%)
Musculoskeletal & Rheumatologic	3 (8%)	14 (9%)
Other	5 (5%)	41 (22%)
<i>Country^a</i>		
United States ^b	17 (45%)	51 (31%)
United Kingdom	4 (10%)	32 (20%)
Netherlands	1 (3%)	15 (9%)
Other	16 (42%)	78 (48%)
<i>Intervention Type^a</i>		
Pharmaceuticals	18 (47%)	68 (42%)
Screening	8 (21%)	19 (12%)
Surgical	8 (21%)	25 (15%)
Medical Procedure	5 (13%)	22 (14%)
Health Education	4 (10%)	13 (8%)
Other	11 (29%)	58 (36%)
<i>Study Funder^a</i>		
Government ^c	21 (55%)	54 (33%)
Pharmaceutical	11 (29%)	46 (28%)
Other ^d	10 (26%)	17 (10%)
None	1 (3%)	21 (13%)
Not Stated	1 (3%)	32 (20%)
<i>Prevention Stage^c</i>		
Primary	7 (18%)	23 (14%)
Secondary	10 (26%)	17 (10%)
Tertiary	21 (55%)	122 (75%)

^aNot mutually exclusive^bP<0.1^cP<0.05^dIncludes foundations, professional member organizations, and health care organizations

Table 2.

Descriptions of the stratifications reported in reviewed studies

Variable	Frequency (n=68)
Number of variables used to for stratification	
Univariate: other patient characteristic	47
Multivariate	15
Univariate: predicted risk *	6
Patient characteristics used for stratification *	
Age	35
Lab/Test/Procedure Results	7
History of Comorbidity(ies)	7
Predicted Risk **	6
Severity of Symptoms/Disease	5
History of Target Disease	5
Gender	5
Other	5
Behavioral/Lifestyle Factors	4
Race	3
Employment status	1
Reported sources of heterogeneity in CEA	
One source reported:	35
baseline risk	16
treatment effect	9
life expectancy	7
costs (productivity, intervention or costs of future clinical events)	3
preferences	0
More than one source reported	17
No source reported	16

* Not mutually exclusive

** Predicted risk is single variable created as a composite of multiple patient variables

Table 3.

Changes in CEA value determinations, with subgroup ratios within stratifications compared to average base-case ratios for *U.S. studies* (n=23 stratifications)

Result	Frequency
Stratification creates a ratio that was more favorable than the average CEA ratio at: *	8
<i>Cost-saving threshold*</i>	0
<i>\$50K/QALY threshold*</i>	7
<i>\$100K/QALY threshold*</i>	2
<i>Dominated threshold (less effective, more costly) *</i>	1
Stratification creates a ratio that was less favorable than the average CEA ratio at: *	10
<i>Cost-saving threshold *</i>	0
<i>\$50K/QALY threshold *</i>	3
<i>\$100K/QALY threshold *</i>	9
<i>Dominated threshold (less effective, more costly) *</i>	3
Stratification does not change value-based decision at any threshold	10

* Not mutually exclusive

Note: Only includes U.S. studies reporting an average population base-case ratio

Table 4.

Characteristics of studies reporting heterogeneity (n=38 studies)

Characteristics	Frequency
<i>Heterogeneity in patient characteristics pre-specified</i>	
Yes- all variables	32
Some variables	2
No variables	4
<i>CEA results presented separately for each pre-specified subgroup</i>	
Yes	38
<i>Term used by author to identify heterogeneity</i>	
No specific term used	14
Subgroup/stratified analysis	9
Sensitivity analysis/scenario analysis	15
<i>Statistical precision reported</i>	
Any (Costs/QALYs/ICER)	9
None	29
<i>Authors discussed different interpretations of the value of the intervention for different subgroups</i>	
Yes	15
No	23
<i>Equity concerned discussed</i>	
Yes	0
No	38