Applied Health Economics and Health Policy (2020) 18:477–489
https://doi.org/10.1007/s40258-019-00545-9

REVIEW ARTICLE

Do Research Groups Align on an Intervention's Value? Concordance of Cost-Effectiveness Findings Between the Institute for Clinical and Economic Review and Other Health System Stakeholders

Matthew Sussman[1] · **Jeffrey C. Yu**[1] · **Joseph Menzin**[1]

Published online: 10 January 2020

Abstract
The Institute for Clinical and Economic Review (ICER) employs fixed cost-effectiveness (CE) thresholds that guide their appraisal of an intervention's long-term economic value. Given ICER's rising influence in the healthcare field, we undertook an assessment of the concordance of ICER's CE findings to the published CE findings from other research groups (i.e., "non-ICER" researchers including life science manufacturers, academics, and government institutions). Disease areas and pharmaceutical interventions for comparison were determined based on ICER evaluations conducted from 1 January 2015 to 31 December 2017. A targeted literature search was conducted for non-ICER CE publications using PubMed. Studies had to be conducted from the US setting, include the same disease characteristics (e.g., disease severity; treatment history), incorporate the same pharmaceutical interventions and comparison groups, and present incremental costs per quality-adjusted life-year (QALY) gained from the healthcare sector or payer perspective. Discordance was measured as the proportion of unique interventions that would have had more favorable valuations (i.e., low, intermediate, high value-for-money) if the CE findings from other research groups had been used for decision making instead of ICER's findings. More favorable valuations were defined as transitioning from low value (as determined by ICER) to intermediate or high value (as determined by other researchers) and from intermediate value (as determined by ICER) to high value (as determined by other researchers). Among the 13 non-ICER studies meeting inclusion criteria, six disease areas and 14 interventions were assessed. Of the 14 interventions, a more favorable valuation would have been recommended for ten therapies if the CE ratios from other research groups had been used for decision making instead of ICER's findings, representing a 71.4% (10/14) discordance rate. Moreover, these discrepancies were found in each of the evaluated disease areas, with the largest number of discordant valuations found in rheumatoid arthritis (five out of six interventions were discordant) followed by one valuation each in multiple sclerosis (one out of three), non-small cell lung cancer (one out of two), multiple myeloma (one out of one), high cholesterol (one out of one), and congestive heart failure (one out of one). Our findings indicate high discordance when comparing ICER's appraisals to the CE findings of non-ICER researchers. To understand the value of new interventions, the totality of evidence on the CE of an intervention—including results from ICER and non-ICER modeling efforts—should be considered when making coverage and reimbursement decisions.

1 Introduction

Stakeholders in the US healthcare market, including health plans and pharmacy benefit managers (PBMs), are seeking innovative ways to curtail rising drug prices. CVS Caremark, a PBM, recently instituted an initiative to increase prescription medication affordability, which will involve the use of findings from the Institute for Clinical and Economic Review (ICER) [1]. As such, CVS Caremark will allow their clients, predominantly health plans, to deny coverage for any drug launched with a cost-effectiveness (CE) ratio over $100,000 per quality-adjusted life-year (QALY) gained [1]. Additionally, the Department of Veteran Affairs (VA) Pharmacy Benefits Management Services office has initiated an arrangement to incorporate ICER reviews into the VA formulary decision-making process [2].

In recent months, ICER's findings may have also influenced the pricing decisions of manufacturers. In 2018, both Regeneron/Sanofi and Amgen announced a substantial

✉ Matthew Sussman
msussman@bhei.com

[1] Boston Health Economics, LLC, 265 Franklin Street, Suite 1101, Boston, MA 02110, USA

Key Points for Decision Makers

In addition to the cost-effectiveness (CE) evaluations conducted by the Institute for Clinical and Economic Review (ICER), other research groups (life science manufacturers, academics, government institutions) perform their own independent CE assessments. This study was conducted to compare ICER's CE findings to those published by other research groups in the USA.

Our findings indicate substantial differences when comparing ICER's appraisals to the CE findings of other research groups.

As CE ratios are increasingly used by health system stakeholders (payers, pharmacy benefit managers) as a means to control rising drug pricing, reliance on a single entity and a single model for decision making may be limiting. The totality of evidence on the CE of an intervention, including results from ICER and non-ICER researchers, should be considered when making coverage and reimbursement decisions.

reduction in the list prices of their PCSK9 inhibitors, which resulted in drug prices that aligned with ICER's value-based CE benchmark [3, 4]. While the exact motivations of the price reduction by Regeneron/Sanofi and Amgen are unknown, it is safe to assume that ICER's findings are relevant to the discussion of appropriate drug pricing and coverage in the USA, and thus are the focus of this brief report.

ICER is a non-profit health technology assessment (HTA) organization that evaluates evidence on the value of medical tests, treatments, and delivery system innovation to determine a drug's value to patients and the larger healthcare system [5]. ICER utilizes its Value Assessment Framework, which is updated every 2 years based on stakeholder feedback and the latest methodological trends in the industry, to determine a drug's short-term affordability as well as its long-term valuation (i.e., value-for-money) [6]. To do so, it commits to a fixed and pre-defined CE threshold of $50,000–$175,000 per QALY gained (US dollars (USD)), which impacts its evaluation of the value of an intervention. ICER considers interventions with a cost per QALY gained of less than $50,000 per QALY gained to represent "high long-term value-for-money" (i.e., cost-effective), while interventions above $175,000 per QALY gained are deemed low long-term value (i.e., not cost-effective). Interventions with a cost per QALY gained between $50,000 and $175,000 per QALY gained are deemed to represent "intermediate long-term value-for-money."

Given national scrutiny on rising drug prices, HTA organizations such as ICER that independently evaluate the clinical and economic value of interventions are primed to impact formulary decision-making processes in the USA. While ICER serves as a valuable source in understanding the CE of therapies, relying on a single entity for decision making may be overly restrictive, as recently opined by the National Pharmaceutical Council in an editorial in the American Journal of Managed Care [7]. In addition to ICER's publicly disseminated CE analyses, life science manufacturers, academia, and government institutions also conduct and publish their own CE analyses of interventions. To the best of our knowledge, a comparison of CE ratios presented by ICER and non-ICER researchers has not yet been documented. By performing a comparative review of published CE analyses, the primary objective of this brief report was to assess the level of alignment in CE assessments of pharmaceutical interventions between ICER and non-ICER researchers. A secondary objective was to construct a case study to help readers better understand the degree to which methodologies may differ between ICER and non-ICER studies, and how to critically review CE studies.

2 Methods

For the primary objective of this study, we compared the CE ratios between ICER and non-ICER researchers for disease areas and pharmaceutical interventions evaluated by ICER during a 3-year period from 1 January 2015 through 31 December 2017. During this time period, ICER evaluated 17 disease areas including two separate evaluations in diabetes, totaling 18 assessments and 76 interventions. Based on the disease areas and interventions evaluated by ICER, we subsequently conducted a literature search to identify comparable, published non-ICER CE studies (Table 1).

The targeted literature search was conducted using PubMed, based on the following search string: {*intervention name*} AND ("cost effectiveness"[title/abstract {tiab}] OR "cost-effectiveness"[tiab] OR "economic evaluation"[tiab] OR "economic"[tiab] OR "economic analysis"[tiab] OR "valuation"[tiab] or "value"[tiab] or "economic value"[tiab]). For completeness, a secondary search was also performed using the Cost-Effectiveness Analysis Registry, which is a database of CE studies aggregated by the Tufts Center for the Evaluation of Value and Risk in Health (CEVR) [8].

Studies were limited to those written in English and published in the same disease area and with the same primary intervention(s) as ICER's evaluations. Studies also had to: (1) be conducted from the US setting; (2) include the same disease characteristics (e.g., disease severity, treatment history); (3) incorporate the same pharmaceutical interventions and comparators; (4) utilize the same model time horizon; (5) present a healthcare sector or payer perspective; and (6)

present an incremental CE ratio in the form of incremental cost per QALY gained. Disease characteristics were decided to be comparable between ICER and non-ICER evaluations based on stage of disease (e.g., mild/moderate/severe disease, heart failure with reduced ejection fraction) and/or baseline clinical characteristics from pivotal clinical trials (e.g., baseline Expanded Disability Status Scale (EDSS) level, baseline New York Heart Association (NYHA) Functional Classification).

Discordance was measured as the proportion of unique interventions that would have had more favorable valuations (i.e., low, intermediate, high value-for-money) if the CE findings from other research groups (i.e., life science manufacturers, academics, or government institutions) had been used for decision making instead of ICER's findings. More specifically, the numerator included the number of interventions that would have been placed in a more favorable valuation category using any of the non-ICER results compared to the ICER results. More favorable valuations were defined as transitioning from low value (as determined by ICER) to intermediate or high value (as determined by other researchers) and from intermediate value (as determined by ICER) to high value (as determined by other researchers); when comparing differences in valuations between ICER and non-ICER, a difference was recorded when *any* of the non-ICER CE ratios for a unique intervention demonstrated a difference in the valuation. The denominator included the total number of unique interventions assessed. All incremental cost per QALY gained estimates were inflated to 2018 USD using the medical care component of the US Bureau of Labor Statistics Consumer Price Index.

For the secondary objective of this study, we constructed a case study to ascertain potential differences in methodology between a non-ICER and an ICER study. We elected to compare ICER's evaluation of calcitonin gene-related peptide (CGRP) inhibitors for the preventive treatment of chronic and episodic migraine [9] to a similar recent evaluation undertaken by the co-authors of this brief report [10]. Doing so afforded us a unique opportunity to compare the finer details of methodologies of two contemporary evaluations. Sussman et al. evaluated the CE of erenumab for the preventive treatment of chronic and episodic migraine, and this study was used as the basis for the comparison [10]. This manufacturer-sponsored study was conducted from the US societal and payer perspectives.

3 Results

3.1 Primary Objective: Level of Discordance

Of the 17 disease areas and 76 interventions evaluated by ICER, our review of the literature for comparable diseases and interventions identified six disease areas and 14 interventions for comparison [4, 11–15]. The disease areas and corresponding interventions included: adalimumab plus methotrexate, adalimumab monotherapy, etanercept monotherapy, infliximab plus methotrexate, rituximab plus methotrexate, and abatacept plus methotrexate in

Table 1 ICER and non-ICER study identification

ICER evaluation disease area	Non-ICER studies identified	Non-ICER studies meeting our study's inclusion criteria
Abuse-deterrent formulations of opioids	1947	0
Atopic dermatitis	12	0
Behavioral health integration	1100	0
Congestive heart failure	297	2
Diabetes	42	0
Extracorporeal membrane oxygenation	575	0
High cholesterol	60	1
Low back pain	3484	0
Multiple myeloma	2246	1
Multiple sclerosis	1405	3
Non-small-cell lung cancer	809	2
Obesity management	1104	0
Obeticholic acid	12	0
Osteoporosis	112	0
Ovarian cancer	45	0
Rheumatoid arthritis	2162	4
Tardive dyskinesia	45	0

ICER Institute for Clinical and Economic Review

rheumatoid arthritis (RA); interferon beta-1a, interferon beta-1b, and glatiramer acetate in multiple sclerosis (MS); pembrolizumab and atezolizumab in non-small-cell lung cancer (NSCLC); carfilzomib in multiple myeloma (MM); evolocumab plus statin in high cholesterol; and sacubitril/valsartan in congestive heart failure (CHF). Our review of the literature identified a total of 25 non-ICER CE studies pertaining to these disease areas and interventions.

Of the 25 non-ICER studies, 13 studies were deemed comparable to those of ICER's, as they utilized the same disease characteristics (e.g., disease severity, treatment history), primary intervention, comparison group, time horizon, and cost per QALY gained measure [16–28]. These 13 studies served as the basis for this analysis, and included the 14 interventions indicated above. The remaining 12 non-ICER studies were excluded for design differences: seven non-ICER studies employed a different time horizon, two non-ICER studies chose an effectiveness outcome other than QALYs (e.g., cost per life-year gained, cost per relapse avoided), two non-ICER studies were excluded for having a different comparison group, and one non-ICER study was excluded for having an incongruent study population (Table 1). Because the 12 non-ICER studies were conducted using incongruent design features, comparisons between ICER and the 12 non-ICER studies were not possible. Among the 13 non-ICER studies that met the inclusion criteria, ten of the studies were sponsored by manufacturers and only three were conducted by academic institutions, none of which disclosed funding from the manufacturers of the interventions studied.

Of the 14 interventions, a more favorable valuation would have been recommended for ten therapies if the CE ratios from other research groups had been used for decision making instead of ICER's findings (Table 2, Fig. 1). This resulted in a 71.4% (10/14) discordance rate. Moreover, these discrepancies were found in each of the evaluated disease areas. For instance, five out of six interventions in RA would have had a more favorable valuation if *any* of the CE findings from non-ICER research groups were used instead, including abatacept + methotrexate (MTX), adalimumab monotherapy, etanercept monotherapy, infliximab + MTX, and rituximab + MTX. Both an ICER and a non-ICER assessment valued adalimumab + MTX the same (low value-for-money), while non-ICER CE findings for infliximab + MTX varied (one non-ICER assessment found the intervention to represent intermediate value and the other found the intervention to represent low value). In MS, ICER and non-ICER valuations for glatiramer acetate and interferon beta-1a were the same (low value-for-money), while non-ICER CE findings for interferon beta-1b varied (one non-ICER assessment found the intervention to represent intermediate value and the other found the intervention to represent low value). In NSCLC, the same valuation was found for atezolizumab (low value-for-money), while the valuation for pembrolizumab was found to be different between ICER (low value) and non-ICER (intermediate value) assessments. Finally, each intervention evaluated in MM (carfilzomib + lenalidomide + dexamethasone), high cholesterol (evolocumab + statin), and CHF (sacubitril/valsartan) would have yielded more favorable valuations had the CE findings from non-ICER research groups been used instead. When considering manufacturer- versus non-manufacturer-sponsored studies, the rates of discordance were 61.5% (8/13) among manufacturer-sponsored studies and 75.0% (3/4) among non-manufacturer-sponsored studies (Table 2).

Table 3 provides details regarding the methodological characteristics of the 19 publications (13 non-ICER and six ICER studies) under evaluation in our review, including the model perspective, population, modeling approach, time horizon, cycle length, health states, utility values, interventions, and drug prices.

3.2 Secondary Objective: Case Study

3.2.1 Overview

The Sussman et al. analysis estimated CE ratios for erenumab 140 mg monthly treatment of chronic and episodic migraine from the US commercial payer perspective [10], while ICER conducted similar analyses from the healthcare sector perspective [9]. In the chronic migraine population, Sussman et al. estimated a CE ratio of $23,079 per QALY gained (representing high value-for-money) compared to ICER's estimate of $90,000 per QALY gained (representing intermediate value-for-money). In the episodic migraine population, Sussman et al. estimated a CE ratio of $180,012 per QALY gained (representing low value-for-money) compared to ICER's estimate of $150,000 per QALY gained (representing intermediate value-for-money). While Sussman et al. found a more favorable valuation in the chronic population, ICER found a more favorable valuation in the episodic population. There are a number of methodological differences in the two models that likely contributed to discrepancies in results, such as the estimation and assignment of the treatment effect (i.e., the underlying model structure) as well as a variety of input values (i.e., transition probabilities, direct medical costs, health state utilities).

3.2.2 Model Structure

In both Sussman et al. [10] and the ICER model, the primary treatment effect included the change in monthly migraine days (MMDs); the manner in which the change in MMDs was estimated and subsequently applied differed in both models. Sussman et al. employed a hybrid, Monte Carlo

Table 2 Difference in valuations between ICER studies and manufacturer and non-manufacturer-sponsored studies

Intervention vs. comparator	ICER study		Non-ICER study			Valuation difference[d]
	CE ratio[a]	Valuation	CE ratio[a]	Valuation	Source	
Rheumatoid arthritis						
Abatacept + MTX vs. MTX	$218,046 [11]	Low	$65,158	Intermediate	Yuan 2010[b] [18]	Yes
			$66,289	Intermediate	Vera-Llonch 2008[b] [19]	
Adalimumab (monotherapy) vs. MTX	$235,648 [11]	Low	$95,635	Intermediate	Spalding 2006[b] [16]	Yes
Adalimumab + MTX vs. MTX	$243,197 [11]	Low	$291,828	Low	Spalding 2006[b] [16]	No
Etanercept (monotherapy) vs. MTX	$224,153 [11]	Low	$134,632	Intermediate	Spalding 2006[b] [16]	Yes
Infliximab + MTX vs. MTX	$212,024 [11]	Low	$614,167	Low	Spalding 2006[b] [16]	Yes
			$61,064	Intermediate	Wong 2002[c] [17]	
Rituximab + MTX vs. MTX	$207,040 [11]	Low	$75,789	Intermediate	Yuan 2010[b] [18]	Yes
Multiple sclerosis						
Glatiramer acetate vs. supportive care	$257,358 [12]	Low	$685,143	Low	Earnshaw 2009[b] [22]	No
			$396,823	Low	Bell 2007[b] [20]	
Interferon beta-1a vs. supportive care	$327,196 [12]	Low	$599,105	Low	Bell 2007[b] [20]	No
Interferon beta-1b	$206,293 [12]	Low	$81,568	Intermediate	Pan 2012[b] [21]	Yes
			$442,873	Low	Bell 2007[b] [20]	
Non-small-cell lung cancer						
Atezolizumab vs. docetaxel	$237,800 [13]	Low	$225,591	Low	Aguiar 2017[c] [23]	No
Pembrolizumab vs. docetaxel	$256,586 [13]	Low	$102,049	Intermediate	Huang 2017[b] [24]	Yes
			$102,885	Intermediate	Aguiar 2017[c] [23]	
Multiple myeloma						
Carfilzomib + LEN-DEX vs. LEN-DEX	$216,972 [14]	Low	$116,655	Intermediate	Jakubowiak 2016[b] [25]	Yes
High cholesterol						
Evolocumab + statin vs. statin	$336,283 [4]	Low	$124,024	Intermediate	Gandra 2016[b] [26]	Yes
Congestive heart failure						
Sacubitril/valsartan vs. ACE inhibitor	$56,695 [15]	Intermediate	$50,127	Intermediate	Gaziano 2016[b] [28]	Yes
			$48,842	High	King 2016[c] [27]	
Proportion of *all* evaluations that would have a more favorable evaluation based on non-ICER CE ratio						71.4%
Proportion of *manufacturer* evaluations that would have a more favorable valuation based on non-ICER CE ratio						61.5%
Proportion of *non-manufacturer* evaluations that would have a more favorable valuation based on non-ICER CE ratio						75.0%

ACE angiotensin converting enzyme, *CE* cost-effectiveness, *CHF* congestive heart failure, *DEX* dexamethasone, *ICER* Institute for Clinical and Economic Review, *LEN* lenalidomide, *MS* multiple sclerosis, *MTX* methotrexate, *NA* not applicable, *NSCLC* non-small-cell lung cancer, *RA* rheumatoid arthritis

[a]CE ratios are reported as incremental cost per QALY gained and have been inflated to 2018 US dollars using the medical care component of the US Bureau of Labor Statistics Consumer Price Index

[b]Manufacturer-sponsored study

[c]Non-manufacturer-sponsored study

[d]More favorable valuations were defined as transitioning from low value (as determined by ICER) to intermediate or high value (as determined by other researchers) and from intermediate value (as determined by ICER) to high value (as determined by other researchers); a difference was recorded when *any* of the non-ICER CE ratios for a unique intervention demonstrated a difference in the valuation

patient simulation and Markov cohort model approach. At model start, each patient's mean number of baseline MMDs was simulated based on the mean (standard deviation [SD]) reported in the clinical trial. Each patient then experienced a change in MMDs, simulated using the mean (SD) change reported in the clinical trial. The baseline number of MMDs was subsequently added to the change in MMDs for each patient to estimate the post-treatment MMDs. Based on an open-label extension study for erenumab 40 mg, in which patients experienced a sustained response to treatment through week 60 of the study, a patient's post-treatment MMD was assumed to be sustained for the entire modeling time horizon, up until death. Costs and utilities were assigned at the end of each cycle based on the number of simulated MMDs.

This approach differed from ICER's approach in which a semi-Markov model was conducted with the following health states: (1) active preventive treatment, (2) no

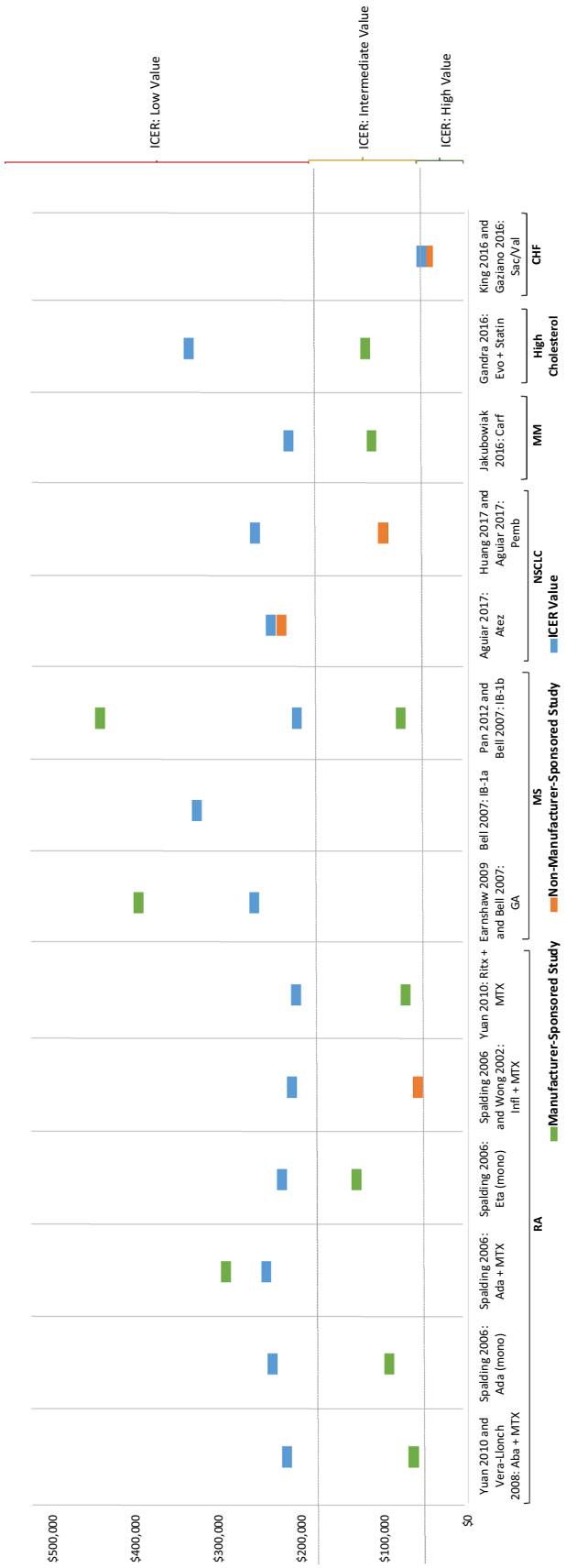

Fig. 1 Comparing CE ratios in ICER studies vs. manufacturer and non-manufacturer-sponsored studies. *CE* cost-effectiveness, *CHF* congestive heart failure, *ICER* Institute for Clinical and Economic Review, *MM* multiple myeloma, *MS* multiple sclerosis, *MTX* methotrexate, *NSCLC* non-small-cell lung cancer, *RA* rheumatoid arthritis, *QALY* quality-adjusted life-year. Incremental cost per QALY gained estimates have been inflated to 2018 US dollars using the medical care component of the US Bureau of Labor Statistics Consumer Price Index

Table 3 Summary table of methodological characteristics in ICER and non-ICER studies

Author and years	Perspective	Population	Modeling approach	Cycle length	Main health states	Utility values	Interventions
Rheumatoid arthritis							
ICER RA Report 2017 [11]	Healthcare sector	Adult patients with moderately to severely active RA who have had an inadequate response to prior therapy	Markov model	6 months	Treatment response at 6 months, responders ACR ≥ 20, non-responders ACR < 20, treatment switch, treatment discontinuation	EQ-5D score = 1 − 1/(1 + exp(2.0734 + 0.0058 × age + 0.0023 × disease duration − 0.2004 × baseline HAQ − 0.2914 × male + 0.0249 × previous DMARDs − 0.8647 × current HAQ))	Adalimumab + MTX Adalimumab (monotherapy) Etanercept (monotherapy) Infliximab + MTX Rituximab + MTX Abatacept + MTX MTX (reference)
Spalding 2006 [9]	Commercial payer	Females aged 55–60 years diagnosed with RA	Markov model	1 year	First-line drug, pooled drugs, death	HUI3 = 0.76 + (0.05 × female) + (0.001 × age) − (0.28 × HAQ)	Adalimumab + MTX Adalimumab (monotherapy) Etanercept (monotherapy) Infliximab + MTX MTX (reference)
Yuan 2010 [11]	Commercial payer	Females aged 55–64 years with moderate-to-severe RA	Monte Carlo simulation model	3 months	Initial HAQ-DI score, expected HAQ-DI score (initial 6 months), expected future HAQ-DI score (beyond initial 6 months)	HAQ-DI 0.00 to < 0.25: 0.86 HAQ-DI 0.25 to < 0.50: 0.80 HAQ-DI 0.50 to < 0.75: 0.76 HAQ-DI 0.75 to < 1.00: 0.71 HAQ-DI 1.00 to < 1.25: 0.66 HAQ-DI 1.25 to < 1.50: 0.59 HAQ-DI 1.50 to < 1.75: 0.51 HAQ-DI 1.75 to < 2.00: 0.43 HAQ-DI 2.00 to < 2.25: 0.33 HAQ-DI 2.25 to < 2.50: 0.23 HAQ-DI 2.50 to < 2.75: 0.12 HAQ-DI 2.75 to 3.00: 0.03	Rituximab + MTX Abatacept + MTX MTX (reference)
Vera-Llonch 2008 [12]	Commercial payer	Females aged 55–64 years with moderate-to-severe RA	Monte Carlo simulation model	3 months	Initial HAQ-DI score, expected HAQ-DI score (initial 6 months), expected future HAQ-DI score (beyond initial 6 months)	Same as in Yuan 2010 [11]	Abatacept + MTX MTX (reference)
Wong 2002 [10]	Commercial payer	Patients with refractory RA	Markov model	6 month	HAQ disability level, death	No disability: 0.836–0.891 Mild disability: 0.714–0.738 Moderate disability: 0.539–0.574 Severe disability: 0.400–0.434	Infliximab + MTX MTX (reference)

Table 3 (continued)

Author and years	Perspective	Population	Modeling approach	Cycle length	Main health states	Utility values	Interventions
Multiple sclerosis							
ICER MS Report 2017 [12]	Healthcare sector	Relapsing-remitting and primary-progressive MS	Markov	1 year	EDSS level, relapse EDSS, death	EDSS 0: – EDSS 1: 0.7905 EDSS 2: 0.7365 EDSS 3: 0.6509 EDSS 4: 0.5816 EDSS 5: 0.5005 EDSS 6: 0.4118 EDSS 7: 0.3000 EDSS 8: −0.0413 EDSS 9: −0.2138	Interferon beta-1a Interferon beta-1b Glatiramer acetate Supportive care (reference)
Bell 2007 [13]	Estimated commercial payer[a]	Patients diagnosed with relapsing-remitting MS	Markov model	1 month	EDSS level, relapse EDSS, death	EDSS 0.0–2.5: 0.824 EDSS 3.0–5.5: 0.679 EDSS 6.0–7.5: 0.533 EDSS 8.0–9.5: 0.491 Utility decrement associated with relapse 0.094	Interferon beta-1a Interferon beta-1b Glatiramer acetate Supportive care (reference)
Pan 2012 [14]	Estimated commercial payer[a]	Patients with MS receiving early treatment	Markov model	6 months	EDSS level, death	EDSS 0–1.5: 0.824 EDSS 2.0–2.5: 0.824 EDSS 3.0–3.5: 0.824 EDSS 4.0–5.5: 0.679 EDSS 6.0–7.5: 0.533 EDSS 8.0–9.5: 0.533	Interferon beta-1b Supportive care (reference)
Earnshaw 2009 [15]	Commercial payer	Patients with relapsing-remitting MS	Markov model	1 month	EDSS level, relapse EDSS, death	EDSS 0.0–2.5: 0.824 EDSS 3.0–5.5: 0.679 EDSS 6.0–7.5: 0.533 EDSS 8.0–9.5: 0.491 Utility decrement associated with relapse: 0.094 Utility decrement associated with PML: 0.000	Glatiramer acetate Supportive care (reference)
Non-small-cell lung cancer							
ICER NSCLC Report 2016 [13]	Healthcare sector	Patients with advanced NSCLC	Partitioned-survival model	1 week	Progression-free, progressive disease, death	1L PF disease, on treatment: 0.78 1L Progressed disease: 0.67 2L PF disease, on treatment: 0.65 2L Progressed disease: 0.47	Atezolizumab Pembrolizumab Docetaxel (reference)
Huang 2017 [17]	Commercial payer	Patients with metastatic NSCLC that expresses high levels of PD-L1	Partitioned-survival model	1 week	Progression-free, progressive disease, death	Time to death (days) ≥360 days: 0.805 (180 days, 360 days): 0.726 (30 days, 180 days): 0.632 <30 days: 0.537	Pembrolizumab Docetaxel (reference)

Table 3 (continued)

Author and years	Perspective	Population	Modeling approach	Cycle length	Main health states	Utility values	Interventions
Aguiar 2017 [16]	Government payer	Patients with NSCLC with or without high PD-L1 expression	Decision-analytic model	1 month	Progression-free disease, post-progression disease, death	Progression-free survival: 0.65 Post-progression survival: 0.43	Atezolizumab Docetaxel (reference for atezolizumab) Pembrolizumab Docetaxel (reference for pembrolizumab)
Multiple myeloma							
ICER Report Multiple Myeloma 2016 [14]	Healthcare sector	Patients with relapsed or refractory multiple myeloma	Partitioned-survival model	1 week	Progression-free disease, progressed disease, dead	Second-line Progression-free disease, on treatment: 0.82 Progression-free disease, off treatment: 0.84 Progressed disease: 0.65	Carfilzomib plus lenalidomide and dexamethasone Lenalidomide and dexamethasone (reference)
Jakubowiak 2016 [18]	Commercial payer	Patients with relapsed multiple myeloma	Partitioned-survival model	4 weeks	Progression-free, post-progression, death	Pre-progression (Cycle 1): KRd 0.810, Rd 0.810 Pre-progression (Cycle 3): KRd 0.818, Rd 0.798 Pre-progression (Cycle 6): KRd 0.829, Rd 0.808 Pre-progression (Cycle 12): KRd 0.840, Rd 0.818 Pre-progression (Cycle 18+): KRd 0.851, Rd 0.829 Post-progression: KRd 0.664, Rd 0.643	Carfilzomib plus lenalidomide and dexamethasone Lenalidomide and dexamethasone (reference)
High cholesterol							
High Cholesterol ICER 2015 [4]	Healthcare sector	Adults with pre-existing ASCVD who require additional lipid lowering	Markov	1 year	Acute incidence cardiovascular event, fatal coronary heart disease or stroke event, non-CVD event, chronic CVD states	No history of cardiovascular disease: 1 History of angina: 0.9064 History of revascularization: 0.9800 History of myocardial infarction (MI): 0.9648 History of MI and revascularization: 0.9818 History of stroke: 0.8835 History of MI and stroke: 0.8524	Evolocumab + statin Statin (reference)
Gandra 2016 [19]	Commercial payer	Patients with HeFH or ASCVD with or without statin intolerance and LDL-C > 100 mg/dL	Markov model	1 year	Cardiovascular disease, established cardiovascular disease, acute coronary syndrome (ACS), ischemic stroke (IS), heart failure (HF), post-ACS, post-IS, post-HF, death	ACS: 0.672 IS: 0.327 HF: 0.602 Post-ACS: 0.824 Post-IS: 0.524 Post-HF: 0.571	Evolocumab + statin Statin (reference)
Congestive heart failure							
ICER Report CHF 2015 [15]	Healthcare sector	Patients with heart failure with reduced ejection fraction	Markov model	1 month	Heart failure hospitalization, additional emergency department visits for heart failure, therapy intolerance, switch therapy, survive, death	Derived from the CHAMPION trial	Sacubitril/valsartan ACE inhibitor (reference)

Table 3 (continued)

Author and years	Perspective	Population	Modeling approach	Cycle length	Main health states	Utility values	Interventions
Gaziano 2016 [21]	Commercial payer	Patients with heart failure with reduced ejection fraction	Markov model	1 month	Heart failure, hospitalized, death	EQ-5D values from mixed-effects model using data from the PARADIGM-HF trial The below are coefficients in the mixed-effects model for EQ-5D: Sacubitril/valsartan: 0.0084 Baseline EQ-5D: 0.5697 Hospitalized within previous 30 days: −0.0771 Baseline age: −0.0011 Time (years): −0.0057 Constant: 0.8991 The below are EQ-5D utilities: Heart failure (sacubitril/valsartan): 0.838 Heart failure (enalapril): 0.829	Sacubitril/valsartan ACE inhibitor (reference)
King 2016 [20]	Commercial payer	Patients with heart failure with reduced ejection fraction	Markov model	3 months	NYHA class, no event, heart failure hospitalization, death	NYHA Class I: 0.815 NYHA Class II: 0.720 NYHA Class III: 0.590 NYHA Class IV: 0.508 HF hospitalization (disutility): 0.100	Sacubitril/valsartan ACE inhibitor (reference)

ACE angiotensin-converting-enzyme inhibitor, *AWP* average wholesale price, *cDMARD* conventional disease-modifying antirheumatic drug, *CHF* congestive heart failure, *CVD* cardiovascular disease, *EDSS* Expanded Disability Status Scale, *EQ-5D* EuroQol-5 Dimensions, *HAQ-DI* Health Assessment Questionnaire-Disability Index, *HUI* Health Utilities Index, *ICER* Institute for Clinical and Economic Review, *IM* intramuscular, *KRd* carfilzomib plus lenalidomide plus dexamethasone, *LDL-C* low-density lipoprotein cholesterol, *MS* multiple sclerosis, *MTX* methotrexate, *N/A* not available, *NSCLC* non-small-cell lung cancer, *NYHA* New York Heart Association, *PML* progressive multifocal leukoencephalopathy, *RA* rheumatoid arthritis, *Rd* lenalidomide plus dexamethasone, *SC* subcutaneous, *TSS* Total Sharp Score, *WAC* wholesale acquisition costs

[a]The base-case analysis was conducted from the societal perspective; data from the publication were manipulated to estimate the commercial payer perspective

preventive treatment, and (3) death. Patients entered the model in the active treatment health state and transitioned to no preventive treatment or death based on rates of discontinuation and US general population mortality. ICER estimated the treatment effect by basing the post-treatment MMDs on the mean baseline MMDs and mean change from baseline in MMDs. While both Sussman et al. [10] and ICER models considered a change in MMDs as the primary treatment effect, the Sussman et al. model did so at the patient level while the ICER model considered the cohort level.

3.2.3 Transition Probabilities

Sussman et al. [10] conducted an indirect treatment comparison of treatment efficacy and discontinuation using Bucher's method [29]. To obtain transition probabilities, ICER conducted a network meta-analysis to estimate the change from baseline in MMDs, days of acute medication use, all-cause discontinuation rates, AE-related discontinuation rates, and AE rates.

3.2.4 Direct Medical Cost Inputs

Sussman et al. [10] applied a drug cost of $6900 to erenumab, which was based on its post-launch list price; ICER, in contrast, applied a drug cost of $5000 to erenumab, which was based on an anticipated pre-launch price. Had ICER applied the post-launch list price of erenumab instead, the CE ratios estimated by ICER would have been even higher than those disseminated, and thus the differences between the two models would have been more pronounced in the chronic migraine population and less pronounced in the episodic migraine population. While Sussman et. al and ICER used different sources for non-pharmacy direct medical costs, the values for costs related to CGRP administration (Sussman et. al: $74.93; ICER: $73.93), emergency department (ED) visits ($948 vs. $949), inpatient hospitalizations ($8954 vs. $8996), and physician office visits ($171 vs. $152) were similar (all in 2017 USD). However, while the non-pharmacy direct medical costs were similar in the Sussman et al. and ICER evaluations, the methods for assigning healthcare resource utilization differed across assessments. In Sussman et al., the probabilities of a physician office visit, ED visit, or hospitalization were assigned based on the number of post-treatment MMDs, while ICER applied treatment efficacy (i.e., percent reduction in MMDs) to baseline rates of primary-care visits, nurse practitioner visits, neurologist visits, other specialist visits, ED visits, and hospitalizations.

3.2.5 Health State Utilities

In Sussman et al. [10], utilities were assessed based on the number of post-treatment MMDs, regardless of migraine attack severity. In contrast, ICER's analysis weighted utilities for chronic and episodic MMDs based on a migraine attack severity distribution. The distribution of mild, moderate, and severe migraines was shifted by a monthly rate based on the severity distribution at the end of a 3-month trial for fremanezumab. At the end of the 3 months, the severity distribution remained the same for the remainder of the model.

4 Discussion

While the results from ICER's analyses are not legally binding with government or commercial payers, ICER has become the de facto HTA agency in the USA, especially in the absence of any other central organization that assists with determining the value of pharmaceutical interventions. It is therefore our belief that ICER should be the focal point of this article. As evidence for this assertion, a recent survey of US payers conducted from 13 August 2018 to 14 February 2019 ($N = 614$ responses) found that more than three-quarters of all surveyed payers had used ICER reports to inform their formulary and reimbursement decisions; in comparison, only 24% of surveyed payers had used the National Comprehensive Cancer Network (NCCN) value assessment framework, followed by the American Society of Clinical Oncology (ASCO) framework (7%), American Heart Association (AHA) framework (5%), and DrugAbacus (< 2%) [30].

Our comparative review of published CE analyses suggests discordance in CE assessments conducted by ICER and non-ICER researchers. In our analyses, we compared CE studies that had the same disease area, population characteristics (e.g., disease severity), comparator groups, and time horizon, as well as presented cost per QALY gained as an outcome. Our findings suggest differences in the valuations of interventions between ICER and non-ICER studies, yielding a 71.4% discordance rate across sponsor types, a 61.5% discordance rate among manufacturer-sponsored studies, and a 75.0% discordance rate among non-manufacturer-sponsored studies. Such high discordance rates were a determining factor for the valuation of an intervention, and in particular, whether an intervention represented high versus intermediate value-for-money or intermediate versus low value-for-money.

In an effort to optimize future CE research and valuations of novel and often expensive interventions, it is important to uncover how ICER reviews differ from those conducted by non-ICER researchers as well as the reasons for the differences. Variations in CE results and valuations between ICER and non-ICER studies are likely driven by methodological differences in the design of the models, as described in the secondary objective of this article. Moreover, while

the discordance rates in this analysis were comparable across manufacturer-sponsored and non-manufacturer-sponsored studies, previously published literature indicates that analyses funded by manufacturers tend to report more favorable results than do those funded by non-manufacturer sources, also potentially leading to differences in CE results and valuations [31–35].

5 Limitations

The primary limitation of our study included identification of a small sample of non-ICER CE models that matched key design features (e.g., outcomes, time horizon) of ICER's models, which in turn limited the number of possible comparisons between non-ICER and ICER studies. While this challenge in finding proper studies for benchmarking highlights a potential need for greater standardization in CE modeling, it also shows the degree of methodological subjectivity inherent in study design. Future research should extend beyond the 3-year identification period used in this study in an attempt to identify a larger pool of comparative non-ICER and ICER studies; doing so may even allow for an analysis of temporal trends between this current evaluation based on the 2015–2017 identification period and a future evaluation based on a 2018–2019 identification period. Additionally, nearly half of the interventions (six out of 14) compared between ICER and non-ICER studies were conducted in RA; thus, a single ICER evaluation may have influenced the results of our analysis.

6 Conclusions

Our findings indicate a high percentage of discordance when comparing ICER's appraisal to the CE findings of other research groups. Based on the CE findings of other research groups, most of the interventions would have received a more favorable valuation, thus shifting from low value-for-money to intermediate value-for-money or from intermediate value-for-money to high value-for-money. Further exploration is required to ascertain the reasons for these discordances, which may be due to differences in model design (e.g., model structure, patient flow, model assumptions, model input values) and/or funding source. We have presented a case study portraying the degree of differentiation that can occur between ICER and non-ICER studies in terms of modeling methodology, and suggest a closer inspection of ICER and non-ICER designs when assessing the value of an intervention.

Health system stakeholders, including payers and PBMs, are increasingly using CE ratios as a means to control drug pricing. Although ICER serves as a valuable source in understanding the value of new interventions, relying on a single entity and a single model for decision making may be limiting. Therefore, the totality of evidence on the CE of an intervention—including results from ICER and non-ICER modeling efforts—should be carefully considered when making coverage and reimbursement decisions.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Matthew Sussman, Jeffrey C. Yu, and Joseph Menzin. The first draft of the manuscript was written by Matthew Sussman and Jeffrey C. Yu and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Compliance with Ethical Standards

Funding This study received no external funding.

Conflict of interest Matthew Sussman, Jeffrey C. Yu, and Joseph Menzin declare no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. CVS Health. Current and new approaches to making drugs more affordable. Updated August 2018. https://cvshealth.com/sites/default/files/cvs-health-current-and-new-approaches-to-making-drugs-more-affordable.pdf. Accessed 16 Feb 2019.
2. Institute for Clinical and Economic Review. The Institute for Clinical and Economic Review to Collaborate with the Department of Veteran Affairs' Pharmacy Benefits Management Services Office. Updated June 27, 2017. https://icer-review.org/announcements/va-release/. Accessed 16 Feb 2019.
3. Robinson JC. Amgen Cuts Repatha's Price By 60% Will Value-based Pricing Support Value-based Patient Access? Health Affairs. Updated November 28, 2018. https://www.healthaffairs.org/do/10.1377/hblog20181127.943927/full/. Accessed 16 Feb 2019.
4. Institute for Clinical and Economic Review. PCSK9 inhibitors for treatment of high cholesterol: effectiveness, value, and value-based price benchmarks. https://icer-review.org/material/high-cholesterol-final-report/. Accessed 20 Nov 2019.
5. Institute for Clinical and Economic Review. About. https://icer-review.org/about/. Accessed 16 Feb 2019.
6. Institute for Clinical and Economic Review. Final value assessment framework for 2017–2019. https://icer-review.org/final-vaf-2017-2019/. Accessed 16 Feb 2019.
7. Dubois, RW. From the Editorial Board: Robert W. Dubois, MD, PhD. Am J Manag Care. 2019;25(7):302. https://www.ajmc.com/journals/issue/2019/2019-vol25-n7/from-the-editorial-board-robert-w-dubois-md-phd. Accessed 15 Aug 2019.
8. Center for the Evaluation of Value and Risk in Health. CEA Registry. https://cevr.tuftsmedicalcenter.org/databases/cea-registry. Accessed 25 Mar 2019.
9. Institute for Clinical and Economic Review. Calcitonin gene-related peptide (CGRP) inhibitors as preventive treatments for patients with episodic or chronic migraine: effectiveness and

value. https://icer-review.org/material/cgrp-final-report/. Accessed 20 Nov 2019.
10. Sussman M, Benner J, Neumann P, Menzin J. Cost-effectiveness analysis of erenumab for the preventive treatment of episodic and chronic migraine: results from the US societal and payer perspectives. Cephalalgia. 2018;38(10):1644–57.
11. Institute for Clinical and Economic Review. Targeted immune modulators for rheumatoid arthritis: effectiveness & value. https://icer-review.org/material/ra-final-report/. Accessed 20 Nov 2019.
12. Institute for Clinical and Economic Review. Disease-modifying therapies for relapsing-remitting and primary-progressive multiple sclerosis: effectiveness & value. https://icer-review.org/material/ms-final-report/. Accessed 20 Nov 2019.
13. Institute for Clinical and Economic Review. Treatment options for advanced non-small cell lung cancer: effectiveness, value, and value-based price benchmarks. https://icer-review.org/material/nsclc-final-report/. Accessed 20 Nov 2019.
14. Institute for Clinical and Economic Review. Treatment options for relapsed or refractory multiple myeloma: effectiveness, value, and value-based price benchmarks. https://icer-review.org/material/mm-final-report/. Accessed 20 Nov 2019.
15. Institute for Clinical and Economic Review. CardioMEMS HF System (St. Jude Medical, Inc.) and sacubitril/valsartan (Entresto, Novartis AG) for management of congestive heart failure: effectiveness, value, and value-based price benchmarks. https://icer-review.org/material/chf_final_report/. Accessed 20 Nov 2019.
16. Spalding JR, Hay J. Cost effectiveness of tumour necrosis factor-α inhibitors as first-line agents in rheumatoid arthritis. Pharmacoeconomics. 2006;24(12):1221–32.
17. Wong JB, Singh G, Kavanaugh A. Estimating the cost-effectiveness of 54 weeks of infliximab for rheumatoid arthritis. Am J Med. 2002;113(5):400–8.
18. Yuan Y, Trivedi D, Maclean R, et al. Indirect cost-effectiveness analyses of abatacept and rituximab in patients with moderate-to-severe rheumatoid arthritis in the United States. J Med Econ. 2010;13(1):33–41.
19. Vera-Llonch M, Massarotti E, Wolfe F, et al. Cost-effectiveness of abatacept in patients with moderately to severely active rheumatoid arthritis and inadequate response to tumor necrosis factor-α antagonists. J Rheumatol. 2008;35(9):1745–53.
20. Bell C, Graham J, Earnshaw S. Cost-effectiveness of four immunomodulatory therapies for relapsing-remitting multiple sclerosis: a Markov model based on long-term clinical data. J Manag Care Pharm. 2007;13(3):245–61.
21. Pan F, Goh JW, Cutter G, et al. Long-term cost-effectiveness model of interferon beta-1b in the early treatment of multiple sclerosis in the United States. Clin Ther. 2012;34(9):1966–76.
22. Earnshaw SR, Graham J, Oleen-Burkey M, et al. Cost effectiveness of glatiramer acetate and natalizumab in relapsing-remitting multiple sclerosis. Appl Health Econ Health Policy. 2009;7(2):91–108.
23. Aguiar PN Jr, Perry LA, Penny-Dimri J, et al. The effect of PD-L1 testing on the cost-effectiveness and economic impact of immune checkpoint inhibitors for the second-line treatment of NSCLC. Ann Oncol. 2017;28(9):2256–63.
24. Huang M, Lou Y, Pellissier J. Cost effectiveness of Pembrolizumab vs. standard-of-care chemotherapy as first-line treatment for metastatic NSCLC that expresses high levels of PD-L1 in the United States. Pharmacoeconomics. 2017;35(8):831–44.
25. Jakubowiak AJ, Campioni M, Benedict Á, et al. Cost-effectiveness of adding carfilzomib to lenalidomide and dexamethasone in relapsed multiple myeloma from a US perspective. J Med Econ. 2016;19(11):1061–74.
26. Gandra SR, Villa G, Fonarow GC, et al. Cost-effectiveness of LDL-C lowering with evolocumab in patients with high cardiovascular risk in the United States. Clin Cardiol. 2016;39(6):313–20.
27. King JB, Shah RU, Bress AP, et al. Cost-effectiveness of sacubitril-valsartan combination therapy compared with enalapril for the treatment of heart failure with reduced ejection fraction. JACC Heart Fail. 2016;4(5):392–402.
28. Gaziano TA, Fonarow GC, Claggett B, Chan WW, et al. Cost-effectiveness analysis of sacubitril/valsartan vs enalapril in patients with heart failure and reduced ejection fraction. JAMA Cardiol. 2016;1(6):666–72.
29. Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997;50:683–91.
30. Science & Innovation Theater. (2019). ICER: payer perspectives on the use and usage of ICER reports [PowerPoint slides]. https://www.amcp.org/Resource-Center/formulary-utilization-management/icer-payer-perspectives-use-and-usage-icer-reports.
31. Bell CM, Urbach DR, Ray JG, et al. Bias in published cost effectiveness studies: systematic review. BMJ. 2006. https://doi.org/10.1136/bmj.38737.607558.80.
32. Norris P, Herxheimer A, Lexchin J, Mansfield P. Drug promotion: what we know, what we have yet to learn. World Health Organization and Health Action International; 2005.
33. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: a systematic review. BMJ. 2003;326:1167–70.
34. Neumann PJ, Sandberg EA, Bell CM, et al. Are pharmaceuticals cost-effective? A review of the evidence. Health Aff. 2000;19(2):92–109.
35. Chopra SS. Industry funding of clinical trials: benefit or bias? JAMA. 2003;290(1):113–4.