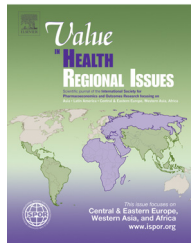
VALUE IN HEALTH REGIONAL ISSUES 4C (2014) 66–74

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/vhri



Cost-effectiveness versus Cost-Utility Analyses: What Are the Motives Behind Using Each and How Do Their Results Differ?—A Polish Example

Joanna Jakubiak-Lasocka, MSc Eng[1], *Michał Jakubczyk, PhD*[2,*]

[1]*HealthQuest spółka z ograniczoną odpowiedzialnością Sp. K., Warsaw, Poland;* [2]*Institute of Econometrics, Warsaw School of Economics, Warsaw, Poland*

ABSTRACT

Objectives: We aimed to compare the use of cost-effectiveness analysis and cost-utility analysis in health technology assessment in Poland. **Methods:** We analyzed all the submissions (155) made to the Polish Agency for Health Technology Assessment in the period 2007 to 2011, with 316 intervention-comparator comparisons reporting incremental cost-effectiveness ratios (ICERs) or incremental cost-utility ratios (ICURs). We compared ICERs and ICURs when both were reported (31%), determined factors associated with reporting one or the other, and tested the precision of their assessment. **Results:** In 13% of the cases, ICER and ICUR led to different decisions (were on opposite sides of the willingness-to-pay threshold). Cost-effectiveness analyses were more frequently performed in oncology, offering at the same time more favorable results. It was also more frequent for longer time-horizon models, although then ICER values were on average higher. **Conclusions:** In Poland, cost-utility analysis is a usual approach of increasing popularity. Interestingly, although assessing ICUR requires additional assumptions, it is estimated more precisely (reported ranges of values in sensitivity analyses are narrower), especially in oncology. ICER and ICUR disagree more often than previously shown in literature. There seem to be no clear signs of biases in submissions (selecting whether to present ICER or ICUR on the basis of their values), but the current study is limited because only the values presented by manufacturers in the submission are available.

Keywords: cost-effectiveness analysis, cost-utility analysis, HTA, incremental cost-effectiveness ratio, incremental cost-utility ratio, Poland.

Introduction

Health technology assessment (HTA) reports presenting the rationale to reimburse a new health technology typically encompass an economic analysis, that is, quantification of additional cost of using this technology in relation to additional health effects [1,2]. Health effects are usually measured as life-years gained (LYGs) or quality-adjusted life-years (QALYs), accounting also for the quality-of-life outcomes [3]. An economic analysis in which LYGs are used is often referred to as cost-effectiveness analysis (CEA) with its parameter of interest being called incremental cost-effectiveness ratio (ICER), whereas an analysis in which QALYs are used is often called cost-utility analysis (CUA) and the resulting parameter is called incremental cost-utility ratio (ICUR). The terms ICER and ICUR are sometimes not distinguished and the context tells whether the result is expressed in LYGs or QALYs. The ICER/ICUR is then compared with the (official or approximate) willingness to pay (WTP) for a unit of effect, that is, threshold to make a final recommendation.

Because HTA aims to evaluate the complete economic and clinical consequences, it would seem natural to favor CUA over CEA. Indeed, although national HTA guidelines differ and may be sometimes vague (cf. Table 1), CUA is overall preferred (strongly preferred in six countries, somewhat preferred in two, not preferred in five). Agency for Health Technology Assessment in Poland (AHTAPol) guidelines treat CEA and CUA equally, demanding that the choice between them be justified; however, the Polish Reimbursement Act, which came into force in 2012, strictly prefers CUA.

This article aims to compare the use of CEA and CUA in HTA in Poland via econometric analysis of data in submissions. In particular, we intend to analyze 1) what leads to the selection of CEA or CUA; 2) how their results differ (point values and range in sensitivity analysis); and 3) whether any bias is present, that is, preferring a more favorable type of analysis to present in a submission.

Our analysis can help to answer the question whether such regulations, obliging manufacturers to present CUAs, were needed in Poland (e.g., when there seems to have been some bias in selecting CEA or CUA) and how they can affect the natural path of HTA development. Such analysis may support other countries in the Central and Eastern European region in shaping their formal HTA regulations.

Conflicts of interest: The authors have indicated that they have no conflicts of interest with regard to the content of this article.

* *Address correspondence to:* Michał Jakubczyk, Institute of Econometrics, Warsaw School of Economics, al. Niepodległości 162, 02-554 Warsaw, Poland.

E-mail: michal.jakubczyk@sgh.waw.pl.

2212-1099$36.00 – see front matter

http://dx.doi.org/10.1016/j.vhri.2014.06.008

Table 1 – Summary of selected national guidelines for the form of economic evaluation.

Country, institution; guidelines, year	Recommendations for the analytic method in economic evaluation	CUA preferred over CEA?
Australia, Pharmaceutical Benefits Advisory Committee; Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee, 2008 [4]	"Cost-utility analysis (generally preferred). (…), a cost-utility analysis is the preferred form of economic evaluation for either or both of the following situations: - where there is a claim of incremental life-years gained in the economic evaluation - in order to assess the impact of quality adjusting that survival gain - where relevant direct randomised trials report results using a MAUI."	Yes, CUA generally preferred.
Belgium, Belgian Health Care Knowledge Centre; Belgian Guidelines for Economic Evaluations and Budget Impact Analyses: Second Edition, 2012 [5]	"Cost-effectiveness analysis should be used if improving life expectancy is the main objective of the treatment (…). Cost-utility analysis should be used if the treatment has an impact on health-related quality of life that is significant to the patient or if there are multiple patient-relevant clinical outcome parameters expressed in different units. If a cost-utility ratio is presented as a reference case analysis result, the corresponding cost per life-year gained should also be presented."	No. CUA preferred if the treatment has an impact on HRQOL. Should be accompanied by CEA.
Canada, Canadian Agency for Drugs and Technologies in Health; Guidelines for the Economic Evaluation of Health Technologies, 2006 [6]	"A CUA should be used in the Reference Case where meaningful HRQL differences between the intervention and alternatives have been demonstrated, and where appropriate preference (utility) data are available. A CEA should be used as the Reference Case when a CUA is an inappropriate choice. Use a final outcome (e.g., life-years gained), or if that is impossible, an important patient outcome. (…)"	Qualified yes.
France, Collège des Économistes de la Santé; French Guidelines for the Economic Evaluation of Heath Care Technologies, 2004 [7]	"Each evaluation has its own particular scope and limitations. The type of study selected should be clearly stated and justified with respect to the issue addressed, and must be described at the start of the study. The author should also provide his personal definition of the type of study used."	No, it depends on the case.
Finland, Ministry of Social Affairs and Health; Guidelines for Preparing a Health Economic Evaluation, 2009 [8]	"The choice of the method of analysis most suitable for each situation (cost-minimization analysis, cost-effectiveness analysis, cost-utility analysis and cost-benefit analysis) depends primarily on how the therapies compared affect patients' health state."	No, it depends on the case.
Ireland, The Health Information and Quality Authority; Guidelines for the Economic Evaluation of Heath Technologies in Ireland, 2010 [9]	"The preferred evaluation type for the reference case is a cost-utility analysis (CUA) with the outcomes expressed in terms of quality-adjusted life-years (QALYs). In exceptional circumstances, a cost-effectiveness analysis (CEA) with the outcomes expressed in terms of life-years gained (or other relevant outcome if the technology does not add life-years) may be used as the reference case when a cost-utility analysis is an unsuitable choice. Clear, detailed empirical evidence must be provided to justify this position."	Yes.
The Netherlands, College voor zorgverzekeringen; Guidelines for Pharmacoeconomic Research, 2006 [10]	"If the improvement in quality of life forms an important effect of the drug being assessed, then it is necessary to carry out a cost-utility analysis (CUA). If this is not the case, then a cost-effectiveness (CEA) has to be carried out. (…)"	No, it depends on the case.
New Zealand, Pharmaceutical Management Agency (PHARMAC); Guidelines for Funding Applications to PHARMAC, 2010 [11]	"Economic analyses should be in the form of a CUA, with benefits measured in terms of quality-adjusted life-years (QALYs). In cases where the clinical outcomes of the drug and the comparator have been shown to be equivalent, a cost-minimisation analysis may be appropriate. Other forms of cost-effectiveness or cost-benefit analyses (CBA) should not be provided to PHARMAC."	Yes.

continued on next page

Poland, Polish Agency for Heath Technology Assessment; Guidelines for Conducting Health Technology Assessment, 2009 [12]	"Analytical method is always selected according to health effects identified and measured and the choice should always be justified. A standard economic analysis as part of a HTA report should be composed of: - cost-consequences analysis, - cost-effectiveness analysis or cost-utility analysis; if there are no differences in clinical effectiveness between health technologies compared the cost-effectiveness analysis may be replaced with cost minimisation analysis. The choice of one method does not exclude using another one as an additional analysis, if the author finds it justified."	No, it depends on the case.
Scotland, Scottish Medicines Consortium; Guidance to Manufacturers for Completion of New Product Assessment Form, 2013 [13]	"In general, cost-utility analysis is the preferred form of economic evaluation, with health effects expressed in terms of quality-adjusted life-years (QALYs)."	Yes.
Spain; Spanish Recommendations on Economic Evaluation of Health Technologies, 2010 [14]	"Develop a cost-effectiveness analysis when there is a clinically significant effect (improvement in health) and sufficient information is not available to perform a cost-utility analysis."	Qualified yes. CEA should be performed when CUA-needed data are lacking.
Sweden, The Pharmaceutical Benefits Board; General Guidelines for Economic Evaluations from the Pharmaceutical Benefits Board, 2003 [15]	"Cost-effectiveness analysis is recommended, with quality-adjusted life-years (QALYs) as the measure of effect. In treatments that mostly affect survival, both QALYs and gained life-years should be shown (…). If it is difficult to use QALYs (e.g. with heavy pain over a short time in connection with treatment), then a cost-benefit analysis with the willingness to pay may be used as a measure of effect."	Yes.
United Kingdom, National Institute for Health and Care Excellence; Guide to the Methods of Technology Appraisal, 2013 [16]	"For the reference case, cost-effectiveness (specifically cost–utility) analysis is the preferred form of economic evaluation. This seeks to establish whether differences in costs between options can be justified in terms of changes in health effects. Health effects should be expressed in terms of QALYs."	Yes.

HRQOL, health-related quality of life; HTA, health technology assessment.

A similar analysis was undertaken by Chapman et al. [17], who compared published results of CEA/CUA. They compared 173 ICER/ICUR pairs (from 63 articles in the CUA database at the Harvard Center for Risk Analysis) published before 1998 to investigate the effect of including health-related quality of life (HRQOL) in the analysis. They concluded that the two approaches differ only slightly and usually lead to the same decision. Research by Tengs [18] and Greenberg and Neumann [19] confirmed these results for oncological interventions, showing a tendency for ICURs to be higher than ICERs (in approximately two-thirds of the situations).

Methods

We follow the general methodology of Chapman et al. [17], applying it to a data set consisting of all manufacturers' submissions to the AHTAPol in the period 2007 to 2011. Therefore, the data set is more recent and homogenous—it relates to a single country and currency and more or less single WTP threshold (possibly increasing over time). The data set includes submissions reporting CUA, CEA, and both, thus enabling detection of the determinants that affect the choice of one form of the analysis and not the other. The use of submissions data reduces the risk of publication bias evident in published articles. In 2012, the Reimbursement Act changed policy to preference for CUA where possible; before this, the analysis reflected the choice of the analyst. Thus, the use of submissions is less regulation-dependent and more transferable to other countries.

Our analysis consists of three parts. First, we compare ICERs and ICURs in reports that include both of them. We quantify the difference between the two and determine how often adjusting for HRQOL affects the decision about allocating resources. Second, we analyze all the data collected and try to determine the factors that influence the choice of ICER or ICUR. Finally, we analyze the differences in estimation precision for ICERs and ICURs to verify whether the precision of ICURs is lower than that of ICERs (under the premise that the quality adjustment added to the calculation process may intuitively increase the uncertainty) and to identify the factors that are associated with the uncertainty of estimation.

Data Sources

We searched through all 201 economic analyses submitted by drug manufacturers to AHTAPol in the period 2007 to 2011. Our inclusion criteria were as follows: 1) CEA or CUA with outcome measured in LYG or QALY (i.e., we excluded cost-minimization analyses and analyses with effects expressed in natural health outcomes); 2) costs measured in Polish zloty (PLN); and 3) conducted sensitivity analysis (e.g., one-way, multiway, and probabilistic). The last criterion resulted in removing two reports only. We included analyses that reported multiple ratios because of more than one comparator or intervention. In such cases, we treated each single comparison as a separate data record. We did not select reports on the basis of their quality. Our goal was to assess the status quo among all the analyses submitted to AHTAPol.

We included 155 reports and 316 intervention-comparator comparisons, with 14 presenting ICER only, 204 ICUR only, and the remaining 98 both. Thus, ICER and ICUR were presented in 35% and 96% of comparisons, respectively.

We extracted the maximum and minimum value of ICER/ICUR from the sensitivity analysis, the year when submission was prepared, time horizon of the model (assumed 99 years in lifelong models), and disease information (a binary variable: 1—intervention used in oncology; 0—all other interventions). We did not explicitly account for the comparator (e.g., do nothing or active treatment) because we concentrated on detecting when CEA/CUA is selected and not on the absolute levels of ICER/ICUR.

We did not include information on the drug manufacturer (55 different companies) or the company preparing the report (14 different commercial or academic entities). We decided not to use them because in our opinion that would require introducing too many dummy variables and make the estimation technically impossible or result in possible spurious conclusions because of multiple hypotheses testing.

The selection of variables was influenced by Chapman et al. [17] and Devlin and Parkin [20]. Instead of interpreting whether the condition or intervention is acute or chronic, however, we decided to collect information about the time horizon of the model, which seems to be more objective and analysis-oriented. We also decided to put special emphasis on oncology because most treatments for cancer are generally expensive and associated with relatively short survival as well as HRQOL.

Because some of the ICER/ICUR values were very large, owing to technical reasons, we introduced the upper limit of 10 million PLN (set arbitrarily, much above the perceived WTP of around 100,000 PLN at that time, €1 $\approx$ 4.1 PLN). We assumed that a dominated technology has ICER/ICUR (the type of analysis—CEA or CUA—was declared in the report) equal to this upper cap—that is, much above the threshold and in effect not recommended for any WTP. Thanks to this substitution, the dominated technology has got a numerical value assigned to it and it is at least as large as any ratio value in the data set. Analogously, a dominant technology has ICER/ICUR equal to 0 (i.e., recommended for any WTP). There were no situations (within our inclusion criteria) in which positive ICER/ICUR resulted from a technology reducing both cost and effects.

Statistical Methods

We used both parametric (Pearson) and nonparametric (Spearman) correlation coefficients. Nonparametric tests were generally used to compare distributions: the Wilcoxon test in the case of paired data and the Mann-Whitney U test in the case of unpaired data. We also present the results of parametric, paired t tests, whenever this yields additional insight. Significance level $P^* = 0.05$ was adopted for these comparisons. While building logistic regression models, potential regressors were removed one by one until all P values were <0.1 (we used a bit higher significance level so as to improve sensitivity in detecting determinants). Because we aimed at identifying determinants and not making forecasts, we used statistical significance criterion and not model fit. We tried to include interaction terms for variables that were themselves statistically significant. Missing values were not imputed and resulted in dropping observations. All the calculations were carried out with R 2.15.0.

Comparison of ICER and ICUR

In this part we restricted data to the subset that presented both the ICER and the ICUR. We determined the frequency of the difference between ratios being positive and negative and calculated the ICUR as a percentage of the ICER, that is, the relative impact of life quality weighting. We compared the distributions using the Mann-Whitney U test. We determined the number of cases when the ICER and the ICUR lay on different sides of the cost-effectiveness threshold of 100,000 PLN, thus leading to different decisions (since 2012 this threshold is defined by law as triple annual gross domestic product per capita: 99,543 PLN as of September 2012, increasing to 105,801 PLN at the beginning of November 2012; before 2012, the threshold was informally perceived to be between 80,000 and 110,000). Finally, we performed a

logistic regression analysis to determine the factors associated with the ICER/ICUR being more favorable. We included as explanatory variables the year the report was made, disease information, and time horizon of the model.

Determinants for Reporting the ICER or the ICUR

To determine factors affecting the choice of the type of analysis, we included all the comparisons, that is, the ICER and the ICUR also reported individually. Because of the strong asymmetry in the data (the ICUR not presented in only 4.4% of the cases and the ICER in 64.6%), we focused on a model explaining the publication of the ICER. We used logistic regression with a year of the performance, disease information, and time horizon of the model as potential regressors. We also used (as a potential regressor) the binary variable denoting whether the ICUR was above or below the WTP threshold (to check whether high evaluations of ICURs led to more frequent ICER reporting). For the completeness of the study, we built an analogous model for the ICUR; however, the current data set hinders strong inferences.

We also included the precision of ICER/ICUR estimation to detect whether, for example, poor precision of one leads to a more frequent reporting of the other. We experimented with three measures of an error. In each, the maximal and minimal value of the ICER/ICUR (as presented in the report in any kind of sensitivity analysis) was used. Equations 1 to 3 present three definitions used for the ICER (analogous formulas were used for the ICUR).

$$\mathrm{ICER}_{\mathrm{error1}} = \mathrm{atan}\left(\mathrm{ICER}_{\max}/100{,}000\right) - \mathrm{atan}\left(\mathrm{ICER}_{\min}/100{,}000\right) \quad (1)$$

$$\mathrm{ICER}_{\mathrm{error2}} = \left(\mathrm{ICER}_{\max} - \mathrm{ICER}_{\min} + 1\right)/\left(\mathrm{ICER} + 1\right) \quad (2)$$

$$\mathrm{ICER}_{\mathrm{error3}} = \left(\mathrm{ICER}_{\max} + 1\right)/\left(\mathrm{ICER}_{\min} + 1\right) \quad (3)$$

The first was defined as an angle in the cost-effectiveness plane between the lines denoting the maximal and minimal ICER/ICUR. The slope coefficients were divided by 100,000 so as the line of WTP was a 45-degree line halving the first quadrant. The second one relates the difference between minimal and maximal value to the base-case value. The third one is a relative measure of difference between minimal/maximal values. We added 1s to denominators so as to avoid the division-by-zero problem.

Precision of the Assessment of the ICER and the ICUR

First, we determined how often maximum and minimum (as taken from sensitivity analysis) corresponding ratios were on opposite sides of the WTP threshold. This is a crude indicator of how precise both measures are, and a direct indicator of whether this imprecision poses a problem to a decision maker. Second, we tried to verify the intuition that including additional HRQOL parameters usually worsens the precision of the estimates, and so ICUR-related errors should be greater than ICER-related errors. We investigated the three measures defined above. We performed a nonparametric Wilcoxon test to compare the precision of assessment between ICERs and ICURs, when presented together and separately. Last, we tried to identify factors associated with differences in ICER/ICUR precision of estimation.

Results

General Data Description

In 11 of 112 CEAs and in 36 of 302 CUAs, the intervention turned out to be dominant and in case of 3 CUAs it was dominated. In other cases, the intervention was more effective and costlier.

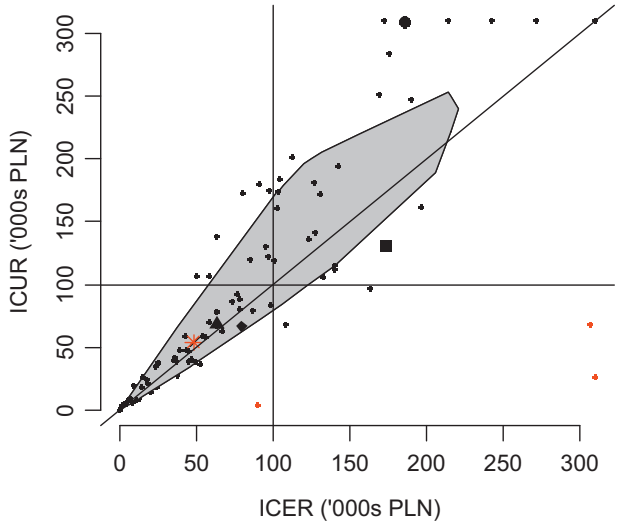


Fig. 1 – Bagplot of ICERs and ICURs; points are drawn when both are available (values in this graph, though not in calculations, are further limited to 310,000 PLN to make it more readable); "*" denotes the bivariate median; "▲"/"♦" denote the pair of univariate medians for paired/all ratios; "■"/"●" denote the pair of univariate means for paired/all ratios; the gray bag contains 50% of all observations. ICER, incremental cost-effectiveness ratio; ICUR, incremental cost-utility ratio; PLN, Polish zloty.

Figure 1 presents the data in the form of a bagplot (including the truncation value of 310,000 PLN for the sake of the graph being readable) along with approximate threshold of 100,000 PLN and 45-degree line.

ICER and ICUR estimations are positively correlated with Pearson linear correlation coefficient of 0.5091 ($P^* < 0.0001$) and Spearman coefficient of 0.9105 ($P^* < 0.0001$). Usually, ICERs/ICURs are below the threshold—the medians for all available values of the ICER and the ICUR amount to 79,389 PLN/LYG and 66,886 PLN/QALY, respectively ("♦"). Calculating the medians for paired values only ("▲") or calculating the bivariate median ("*") does not change the results qualitatively.

Mean ICERs/ICURs calculated when both were available equal 173,775 PLN/LYG and 130,606 PLN/QALY ("■"). Thus, the ICER is greater on average. The majority of data points, however, lie above the 45-degree line, meaning that usually, when both are reported, the ICUR is greater (thus including HRQOL in the analysis decreases the cost-effectiveness of the technology being studied).

Calculating means for all (nontruncated) the data yields 185,606 PLN/LYG and 309,000 PLN/QALY ("●"). The increase in both values means that both ratios are greater on average when unaccompanied. The increase is much larger for the ICUR (and also more reliable because of the larger number of cases when the ICER is not reported).

Intuitively, there is a positive skew in the data for the ICER and the ICUR (means larger than medians). Highest (noncapped) ratios amount to 4,606,000 PLN/LYG and 17,430,000 PLN/QALY. This suggests using nonparametric tests when comparing values for both types of ratio (and capping). The first quartiles amount to 24,060 PLN/LYG and 18,420 PLN/QALY, whereas third quartiles amount to 164,500 PLN/LYG and 144,700 PLN/QALY.

Comparison of Paired Ratios

The previous section showed that ICERs and ICURs have similar values—the difference between unpaired ICERs and ICURs is not

statistically significant (Mann-Whitney *U* test, $P^* = 0.3512$). Still, the differences between the two, limited to comparisons when both are reported, present some pattern. The ICER was greater than the ICUR in only 28.5% of the cases (28 of 98) and smaller in 59.2% (58 of 98). In the remaining 12.2%, both were equal (to 0 because interventions were dominant). Therefore, usually including HRQOL in the analysis reduces the additional effect of the technology being evaluated. The paired nonparametric test shows this qualitative difference to be statistically significant (Wilcoxon paired test, $P^* = 0.0049$).

The analysis of means yields a different picture. As mentioned above, the average ICER is greater by 43,169 PLN in cases in which both the ICER and the ICUR are presented, though the difference is not statistically significant (*t* test, $P^* = 0.3471$). Therefore, in the less frequent situations when including HRQOL reduces the cost-effectiveness ratio, this reduction can be substantial. Chapman et al. [17] in their study obtained quite the opposite (Cf. Discussion).

In 35% of the cases, quality of life adjusting changed the cost-effectiveness ratio by less than 20% (i.e., the ICUR was within ±20% bounds of the ICER), and in 70% of the cases by less than 40%. The maximal relative percentage difference amounted to approximately 120%.

In 33 cases (34%) ICERs and in 38 cases (39%) ICURs were above 100,000 PLN, assumed approximately as the threshold value in Poland. In 13 cases (13%), the two were on the opposite sides of the threshold, and thus pointed to different resource allocation decisions. In 4 cases, the ICUR was favorable (while the ICER was unfavorable) and in 9 cases vice versa. This again somewhat differs from Chapman's et al. [17] study, in which 8.1% of the cases were on the other side of the $50,000 threshold (and only 6.4% for the threshold equal to $100,000).

Table 2 – The final logistic regression modeling results for the ICER being greater than the ICUR (the ICUR being more favorable for the technology).

Explanatory variable	Estimate	Odds	P
Intercept	823.1925	NA	0.0790
Year of submission	−0.4103	0.66	0.0786
Oncology (1 = yes)	−1.3401	0.26	0.0415
Time horizon (y)	0.0171	1.02	0.0229

ICER, incremental cost-effectiveness ratio; ICUR, incremental cost-utility ratio; NA, not applicable/available.

Finally, we performed a logistic regression analysis to identify determinants of ICER > ICUR (of introducing HRQOL positively affecting the technology being assessed). In the final model, we retained three variables—the year of performance, time horizon of the model, and whether the intervention was used for oncology—that proved to be significant (Table 2). All the remaining variables and interaction terms between variables from Table 2 were insignificant with $P^* > 0.1$.

Determinants for Reporting ICERs or ICURs

Fig. 2 visualizes the distributions of ratios when reported separately and paired (there are only 14 cases in which the ICER was unaccompanied by the ICUR). The data suggest that unpaired ICER is usually greater and this is confirmed by a nonparametric test (Mann-Whitney *U* test, $P^* < 0.0001$). In the case of the ICUR, the nonparametric approach yields no difference ($P^* = 0.9293$), although the parametric one does (*t* test,

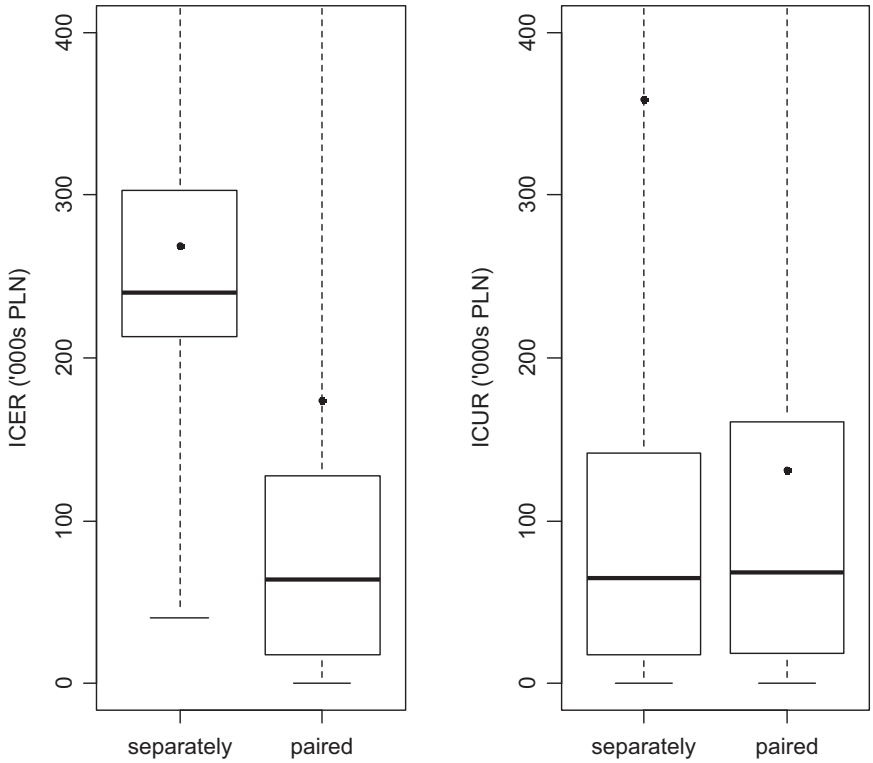


Fig. 2 – Box plot of ICERs and ICURs when paired vs when presented separately (values in this graph, though not in calculations, are further limited to 400,000 PLN to make it more readable); the line denotes the median, the box denotes quartiles, and the whiskers denote min and max; the dot represents the mean. ICER, incremental cost-effectiveness ratio; ICUR, incremental cost-utility ratio; PLN, Polish zloty.

Table 3 – Logistic regression modeling results for factors associated with ICER/ICUR reporting.

Explanatory variable	Estimate	Odds	P
Determinants of the ICUR being published			
Intercept	−2043.6523	NA	0.0004
Year of submission	1.0193	2.77	0.0004
Oncology	−1.7927	0.17	0.0027
Determinants of the ICUR being published (alternative model)			
Intercept	−2679.4330	NA	0.0007
Year of submission	1.3359	3.80	0.0007
Is the ICER > 100,000?	−2.9656	0.52	0.0013
Determinants of the ICER being published			
Intercept	−1.8159	NA	<0.0001
Time horizon (y)	0.0144	1.01	<0.0001
Oncology	2.4113	11.14	<0.0001

ICER, incremental cost-effectiveness ratio; ICUR, incremental cost-utility ratio; NA, not applicable/available.

Table 4 – Logistic regression modeling including precision of estimation as explanatory variables.

Explanatory variable	Estimate	Odds	P
Determinants of the ICUR being published			
Intercept	−1805	NA	0.0319
Year of submission	0.8985	2.46	0.032
Time horizon (y)	−0.024	0.98	0.0889
ICER error 1	0.0823	1.09	0.0023
Determinants of the ICER being published			
Intercept	−1.735	NA	<0.0001
Time horizon (y)	0.014	1.01	0.0004
Oncology	2.534	12.60	<0.0001
ICUR error 3	−0.0002	0.99	0.0081

ICER, incremental cost-effectiveness ratio; ICUR, incremental cost-utility ratio; NA, not applicable/available.

$P^* = 0.02944$), due to a few very high ICUR values. Because of the strong asymmetry, the nonparametric result seems to be more reliable.

We performed the logistic regression analysis to determine factors associated with the reporting of the ICUR. We found two, non-nested models interesting (Table 3). In the first one, presenting the ICUR was explained by the year of submission (reporting increasing in time) and oncology (less frequent reporting in oncology—as explained previously, in oncology ICURs are less favorable than ICERs). In the second model, we introduced a binary variable representing the ICER value being above the WTP threshold (then the oncology loses its significance). High ICER values discourage authors from publishing ICURs. These two models are in fact quite similar because in oncology one can expect high ratios of cost-effectiveness in the first place. The limitation of this part of the analysis is a small sample size (ICURs are usually presented and so there are only a few 0 values in the explanatory variable).

Analogously, we constructed a model explaining the publication of the ICER, which is also presented in Table 3. Using the ICER is more frequent among models with a longer time horizon (however, a longer time horizon is associated with higher values of ICERs, when both ratios are reported). This may be because in lifelong models it is natural to present cost-effectiveness measures stemming solely from life expectancy. The reporting of the ICER is also more popular in oncology (ICURs are less favorable there). ICURs exceeding the 100,000 PLN thresholds did not additionally enhance the odds to present the ICER. No interaction terms were statistically significant.

In addition, we performed a logistic regression analysis including the precision of estimation as an explanatory variable to determine how the uncertainty of estimation affected the decision on ICER/ICUR reporting. Table 4 presents the final results. We experimented with various definitions of precision of estimation, with the one included in the table found to be the most significant. The sign of this association is reverted between models; for example, the large uncertainty related to the ICER is associated with the ICUR also being calculated and the large uncertainty of the ICUR leads to less frequent reporting of the ICER.

Including the precision of estimates resulted in the time horizon of the analysis to be a statistically significant predictor of presenting both the ICER and the ICUR. The longer the time horizon, the greater (lower) the probability of presenting the ICER (ICUR).

Precision of the Assessment of the ICER and the ICUR

We observed a high level of uncertainty of ratio assessment in sensitivity analyses. In 32 of 80 (40%) and 133 of 297 (45%) cases, $ICER_{max}$ and $ICUR_{max}$, respectively, were above the threshold of 100,000 PLN, whereas their corresponding $ICER_{min}$ and $ICUR_{min}$ were below. Thus, in almost 50% of the cases, a decision maker may feel unsure about the base scenario result and may prefer to ignore the results, focusing on other types of analyses presented in HTA reports.

In most of the cases, the ICER is assessed with greater uncertainty (see Fig. 3 in Supplemental Materials found at http://dx.doi.org/10.1016/j.vhri.2014.06.008). For example, on average, the error defined by Equation 1 is larger for the ICER than for the ICUR by 4.7 degrees (t test paired, $P^* < 0.0001$). Similar results are obtained for errors defined by Equations 2 and 3. Obviously, errors for ICER and ICUR are correlated; for example, Pearson correlation coefficient for error type 1 amounts to 0.9638 and is even greater for error 2 and error 3.

We tried to identify situations in which the ICUR is estimated with greater error (Table 5). No interactions were statistically significant. In time ICUR errors are getting relatively smaller as compared to ICER errors. In oncology, uncertainty related to the ICUR is also smaller.

The second observation may be explained by the fact that low HRQOL in such diseases reduces the impact of uncertainty associated with the effect of treatment on life prolongation, while the uncertainty related to the HRQOL estimation is less meaningful due to short life expectancy. More formally, because $\Delta QALY \approx \Delta LYG \times HRQOL + \text{life expectancy} \times \Delta HRQOL$, the uncertainty of $\Delta QALY$ is small, as the uncertainty of ΔLYG is multiplied by low HRQOL and the uncertainty of $\Delta HRQOL$ is multiplied by low life expectancy.

Table 5 – Logistic regression modeling results comparing the precision of CEA and CUA.

Explanatory variable	Estimate	Odds	P
Determinants of the ICUR error > the ICER error			
Intercept	1184.6765	NA	0.0418
Year of submission	−0.5902	0.55	0.0417
Oncology	−1.7383	0.18	0.0445

CEA, cost-effectiveness analysis; CUA, cost-utility analysis; NA, not applicable/available.

Discussion

Although CUA still raises some controversies and is more resource- and time-consuming than CEA, it is often recommended in national guidelines as a first-choice analysis. We collected all the ICERs and ICURs that satisfied our inclusion criteria from manufacturers' submissions to AHTAPol and analyzed their values, precision of estimation, and factors associated with their presentation.

We find the following facts to be the most interesting. First, even though HRQOL in general provides additional mechanism transmitting treatment benefits, it does not increase cost-effectiveness in Poland (ICURs are usually higher, Fig. 1). The reason is that including HRQOL also decreases the benefits of life prolongation due to possibly lower baseline utility and possible improvements in the HRQOL do not moderate this effect. This is the case especially in oncological diseases (Table 2), which is quite intuitive accounting for the low utility of life.

As time passes CUAs tend to be less favorable than CEAs (Table 2). Because of the short time horizon of our study, it results not from a change in the health technologies assessed (i.e., decreasing possibility to improve HRQOL by newly developed drugs). More probably, as authors are increasingly expected to perform CUA, it is also performed and reported in situations in which it is not favorable. This is confirmed by the analysis of ICUR reporting (Table 3)—ICURs are being reported very often and this trend was increasing in the analyzed time frame. This tendency observed in the period 2007 to 2011 came ahead of the new Reimbursement Act demanding the ICUR whenever possible. The new act also demands a systematic review of utility values to be used in economic analyses. This may strengthen the above tendency of ICURs becoming less favorable and would call for a dedicated study when new data have been gathered. These observations, however, may be biased by the fact that ICERs are not always presented.

Second, there are patterns regarding ICER/ICUR reporting. Because ICURs are usually less favorable, it is not surprising that when the ICER is above the threshold, ICURs are less frequently reported (Table 3, the alternative model). One possible interpretation might be that in case of unfavorable results, researches do not bother collecting utility data and preparing a cost-utility model that is most probably not going to change the overall conclusion (or they do bother, but do not present additional, unfavorable results). ICURs are less frequently shown in oncology, whereas for ICERs it is vice versa (Table 3), which may be considered a bias because ICURs are less favorable in oncology (cf. Table 2; similar findings presented by Tengs [18] and Greenberg and Neumann [19]). However, the long time horizon of an analysis promotes presenting ICERs (Table 3), which is less favorable then (Table 2). Perhaps in lifelong models LYGs are a natural measure of an effect. This also shows that including HRQOL in the model has a larger positive effect on the cost-effectiveness of an assessed technology for longer time horizons because there is time for improvements in utility to accumulate.

It is difficult to interpret the impact of the ICER (ICUR) precision of estimate on reporting the ICUR (ICER). Greater precision in ICER estimation reduces the probability of ICUR reporting, whereas greater precision in ICUR estimation increases the probability of ICER reporting (Table 4). Because the ICUR is omitted in only a few cases, one possible interpretation is that whenever authors do not bother with presenting the ICUR, they also do not perform a very thorough sensitivity analysis, thus reducing the range for ICER values presented. The negative association between ICUR estimation error and the ICER being reported is more difficult to interpret. Perhaps when the CUA is performed and equipped with extensive sensitivity analysis triggering large uncertainty, authors do not bother with presenting the results of simple CEA.

Third, even though estimating the ICUR requires additional parameters in the model, the resulting uncertainty is smaller. Perhaps there is a large uncertainty regarding the life expectancy that is reduced in the CUA by multiplying it by health states utility, while the additional uncertainty stemming from HRQOL estimation is smaller than this reduction. This is supported by the fact that ICURs are relatively more precisely estimated in oncology (Table 5). This finding is more robust to the limitations of our study (i.e., publication bias) because ICURs are available in most of the comparisons, and we would not expect authors to omit the ICER when the related uncertainty is small. ICURs are getting relatively more and more precisely estimated in time (Table 5). The strong preference for the ICUR instead of the ICER enforced in Poland might however change the situation because ICURs would also have to be calculated in situations in which it is difficult to credibly assess the HRQOL or its changes.

Interestingly, in case of our data set, the conclusions are quite different from those of Chapman et al. [17] (see Table 6 in Supplemental Materials found at http://dx.doi.org/10.1016/j.vhri.2014.06.008). In the study by Chapman et al., ICURs were usually more favorable than ICERs. This is a qualitative difference from our study. The most plausible explanation is the lack of ICUR reporting bias in Poland, leading to a frequent publishing of nonfavorable ones. There are also qualitative differences regarding the relative values of ICERs and ICURs (that are reversed in our study than in the study by Chapman et al.), but that is not that important in our opinion because the mean values may be driven by individual very high values (large asymmetry in both our and Chapman et al.'s data set).

Because the aim of economic evaluations is to support the decision making in health care, one of the most important conclusions is that in the case of Chapman only 8% of paired ratios were on the opposite sides of the threshold, affecting the resource allocation decision, whereas in case of Polish data it is 13%. Although the numbers are not that different, we would say that this percentage is significant because it emphasizes how important it is to select the right approach. As we collected minimum and maximum values in sensitivity analyses, we additionally noticed that in almost 50% of the cases the extreme values were on the opposite sides of the WTP threshold, which forces a decision maker to decide whether to adopt a conservative approach or a base-case scenario.

One important limitation of our study is the lack of possibility to observe the values that are omitted by the authors of the submissions. It is therefore difficult to discover the true data-generating process, that is, the process that led, for example, to ICER omission. Currently, according to the new Reimbursement Act, AHTAPol requires companies to also submit the implemented model that allows to redo all the calculations and change basic parameters. This means that it will be in principle possible to add CEA to submitted CUA (e.g., by substituting trivial utility values).

In conclusion, the above analysis is not to say that ICURs should be preferred over ICERs in the HTA process or vice versa. The decision should rather be based on economic arguments regarding which parameter describes decision makers' preferences more closely (and that reasoning would most likely point to ICURs). We can, however, say that CUA had been very popular in Poland before the new regulation came into force and that regulation actually sanctioned HTA practice. No clear signals of biases (selective reporting of the ICUR or the ICER) can be seen in the data (though the data-generating process remains hidden because of the limitations mentioned above). Thus, the regulation was not actually needed in Poland. Still, it is worth noting that Poland is quite mature a market regarding the HTA process,

which may explain why HTA practice preceded the regulation. In less HTA-experienced countries, it might make sense to force using CUA by law, but that would most likely make it more difficult for oncology-related technologies to prove cost-effective. Reassuringly, introducing additional parameters into modeling (utilities) does not seem to worsen the precision of the estimates. Luckily then, it seems that we may not have to choose between being "roughly right" or "precisely wrong" but may be a little bit more precisely right in HTA with CUA.

Acknowledgments

This study was made possible by the Polish Agency for Health Technology Assessment (AHTAPol) sharing its data and that is greatly acknowledged. We stress that the data on drugs' manufacturers and HTA reports' authors were anonymized. At the time of data collection and paper writing, the first author worked at AHTAPol. We also thank Elisabeth Fenwick for her valuable comments. The authors have no conflict of interest.

Source of financial support: The authors have no other financial relationships to disclose.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article). http://dx.doi.org/10.1016/j.vhri.2014.06.008

REFERENCES

[1] Drummond MF, Sculpher MJ, Torrance GW, et al. Economic Evaluation of Health Care Programs. (3rd ed.) New York, NY: Oxford University Press, 2005.
[2] Gold MR, Siegel JE, Russel LB, et al. Cost-Effectiveness in Health and Medicine. New York, NY: Oxford University Press, 1996.
[3] Weinstein MC, Torrance G, McGuire A. QALYs: the basics. Value Health 2009;12(Supp. 1):S5–9.
[4] Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee (Version 4.3). Canberra, Australia: Pharmaceutical Benefits Advisory Committee, 2008. Available from: http://www.ispor.org/peguidelines/source/Australia-Guidelines-for-preparing-submissions-to-the-Pharmaceutical-Benefits-Advisory-Committee-2008.pdf.[Accessed December 8, 2012].
[5] Leemput I, Neyt M, Van de Sande S, et al. Belgian Guidelines for Economic Evaluations and Budget Impact Analyses: Second Edition. Health Technology Assessment (HTA). Brussels, Belgium: Belgian Health Care Knowledge Centre (KCE), 2012: Available from: https://kce.fgov.be/sites/default/files/page_documents/KCE_183C_economic_evaluations_second_edition_0.pdf: (KCE Report 183C. D/2012/10.273/54. [Accessed December 8, 2013]).
[6] Guidelines for the Economic Evaluation of Health Technologies (3rd ed.). Ottawa, Canada: Canadian Agency for Drugs and Technologies in Health, 2006. Available from: http://cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf. [Accessed December 8, 2013].
[7] French Guidelines for the Economic Evaluation of Heath Care Technologies. Paris, France: Collège des Économistes de la Santé, 2004. Available from: http://www.ces-asso.org/docs/France_Guidelines_HE_Evaluation.PDF. [Accessed December 8, 2013].
[8] Guidelines for preparing a health economic evaluation. Decree of the Ministry of Social Affairs and Health, Helsinki. 2009. Available from: http://www.stm.fi/c/document_library/get_file?folderId=71837&name=DLFE-10037.pdf. [Accessed December 8, 2013].
[9] Guidelines for the Economic Evaluation of Health Technologies in Ireland. Cork, Ireland: The Health Information and Quality Authority, 2010. Available from: http://www.hiqa.ie/system/files/HTA_Economic_Guidelines_2010.pdf. [Accessed December 8, 2013].
[10] Guidelines for Pharmacoeconomic Research. Diemen, The Netherlands: College voor zorgverzekeringen, 2006. Available from: http://www.ispor.org/peguidelines/source/HTAGuidelinesNLupdated2006.pdf. [Accessed November 8, 2013].
[11] Guidelines for Funding Applications to PHARMAC. Wellington, New Zealand: Pharmaceutical Management Agency (PHARMAC), 2010. Available from: http://www.pharmac.health.nz/ckeditor_assets/attachments/128/guidelines_for_making_funding_applications.pdf. [Accessed December 8, 2013].
[12] Guidelines for Conducting Health Technology Assessment. Warsaw, Poland: Polish Agency for Heath Technology Assessment, 2009. Available from: http://www.aotm.gov.pl/assets/files/wytyczne_hta/2009/Guidelines_HTA_eng_MS_29062009.pdf. [Accessed December 8, 2013].
[13] Guidance to Manufacturers for Completion of New Product Assessment Form. Glasgow, Scotland: Scottish Medicines Consortium, 2013. Available from: http://www.scottishmedicines.org.uk/Submission_Process/Submission_Guidance_and_Templates_for_Industry/Templates-Guidance-for-Submission/Templates-Guidance-for-Submission. [Accessed December 8, 2013].
[14] Lopez-Bastida J, Oliva J, Antonanzas F, et al. Spanish recommendations on economic evaluation of health technologies. Eur J Health Econ 2010;11:513–20.
[15] General Guidelines for Economic Evaluations from the Pharmaceutical Benefits Board. Solna, Sweden: The Pharmaceutical Benefits Board, 2003. Available from: http://www.ispor.org/peguidelines/source/Guidelines_in_Sweden.pdf. [Accessed December 8, 2013].
[16] Guide to the Methods of Technology Appraisal. London, UK: National Institute for Health and Care Excellence, 2013. Available from: http://www.nice.org.uk/media/D45/1E/GuideToMethodsTechnologyAppraisal2013.pdf. [Accessed December 8, 2013].
[17] Chapman RH, Berger M, Weinstein MC, et al. When does quality-adjusting life-years matter in cost-effectiveness analysis? Health Econ 2004;13:429–36.
[18] Tengs TO. Cost-effectiveness versus cost–utility analysis of interventions for cancer: does adjusting for health-related quality of life really matter? Value Health 2004;7:70–8.
[19] Greenberg D, Neumann PJ. Does adjusting for health-related quality of life matter in economic evaluations of cancer-related interventions? Expert Rev Pharmacoecon Outcomes Res 2011;11:113–9.
[20] Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Econ 2004;13:437–52.