RESEARCH

Meta-analysis of economic evaluation studies: data harmonisation and methodological issues

Bhavani Shankara Bagepally[1,2], Usa Chaikledkaew[1,3], Nathorn Chaiyakunapruk[4], John Attia[5] and Ammarin Thakkinstian[1,6*]

Abstract

Background: In the context of ever-growing health expenditure and limited resources, economic evaluations aid in making evidence-informed policy decisions. Cost-utility analysis (CUA) is often used, and CUA data synthesis is also desirable, but methodological issues are challenged. Hence, we aim to provide a step-by-step process to prepare the CUA data for meta-analysis.

Methods: Data harmonisation methods were constructed specifically considering CUA methodology, including inconsistent reports, economic parameters, heterogeneity (i.e., country's income, time horizon, perspective, modelling approaches, currency, willingness to pay). An incremental net benefit (INB) and its variance were estimated and pooled across studies using a basic meta-analysis by COMER.

Results: Five scenarios show how to obtain INB and variance with various reported data: Study reports the mean and variance (Scenario 1) or 95% confidence interval (Scenario 2) of ΔC, ΔE, and ICER for INB/variance calculations. Scenario 3: ΔC, ΔE, and variances are available, but not for the ICER; a Monte Carlo was used to simulate ΔC and ΔE data, variance and covariance can be then estimated leading INB calculation. Scenario-4: Only the CE plane was available, ΔC and ΔE data can be extracted; means of ΔC, ΔE, and variance/covariance can be estimated accordingly, leading to INB/variance estimates. Scenario-5: Only mean cost/outcomes and ICER are available but not for variance and the CE-plane. A variance INB can be borrowed from other studies which are similar characteristics, including country income, ICERs, intervention-comparator, time period, country region, and model type and inputs (i.e., discounting, time horizon).

Conclusion: Out data harmonisation and meta-analytic methods should be useful for researchers for the synthesis of economic evidence to aid policymakers in decision making.

Keywords: Economic evaluation, CUA, Cost-effectiveness

Background

In the context of ever-growing health expenditure and limited resources, identifying healthcare services that yield the highest benefit at the lowest cost is a priority. Economic evaluation studies (EES) provide a framework to systematize both clinical and economic outcomes [1] Cost-utility analysis (CUA) is commonly applied to compare clinical and economic outcomes by estimating an

*Correspondence: ammarin.tha@mahidol.edu
[1] Mahidol University Health Technology Assessment (MUHTA) Graduate Program, Mahidol University, Bangkok, Thailand
Full list of author information is available at the end of the article

Bagepally *et al. BMC Health Services Research* (2022) 22:202

Page 2 of 10

incremental cost-effectiveness ratio (ICER). The costs are usually measured in a specific country currency, while the health benefit is usually measured as a quality adjusted life year (QALY), i.e., the product of years lived and health utility score ranging from 0 (death) to 1 (perfect health), or disability adjusted life years (DALY) [2, 3]. The ICER, $(\text{Cost}_{\text{intervention}}-\text{Cost}_{\text{comparator}})/(\text{QALY}_{\text{intervention}}-\text{QALY}_{\text{comparator}})$, is under the willingness to pay (WTP) threshold (measured in monetary cost per QALY gained), the health intervention is considered to be cost-effective[4]. The guidelines from Joanna Briggs Institute, Cochrane group [5, 6], mainly provide guidelines towards qualitative synthesis or only systematic review of all sorts of economic evaluation (e.g., cost–benefit analysis, cost minimisation analysis, cost-effective analysis, cost-utility analysis) [7–11]. However, these guidelines have a limited focus on data extraction and data harmonisation process to prepare the data for the meta-analysis [7–11].

Further, many methodological issues in the data synthesis of EESs are more challenging than clinical studies because there are many sources of heterogeneity, including study characteristics (e.g., setting, WTP, country, country income), methodology (time horizon, perspective, data source, model type, input parameters, and assumptions) [8]. This is perhaps why most previous systematic reviews of EESs have performed only descriptive analyses and reported only qualitative findings without applying a meta-analysis (MA) to estimate pooled effect measures.

Although Crespo et al. [8] have described a MA for pooling EES (known as the COMparative Efficiency Research, COMER), it has yet been widely adopted as such MA for clinical outcomes. This might be due to EESs being too heterogeneous to pool or choosing the lesser-known parameter "incremental net benefit" (INB) as the effect measure rather than the more commonly used ICER. However, we believe the choice for pooling INB was justified due to the limitations of the ICER [12]. For instance, a negative ICER may indicate a lower cost compared with higher effectiveness or higher costs along with lower effectiveness of interventions, thus introducing ambiguity in interpretation [8, 13]. In contrast, positive and negative INBs directly indicate cost-effectiveness and non-cost-effectiveness of interventions, respectively, which is the information required by policymakers [14, 15].

Furthermore, the COMER method mainly focused on the statistical methods for pooling but did not suggest a detailed step-by-step method of data extraction and data harmonisation as for various styles and suboptimal quality for reporting of economic evaluations [16]. In addition, assessing heterogeneity and exploring the source of it and publication bias have not been described.

Therefore, our primary focus in this manuscript is to provide a methodological approach for meta-analysis of cost-utility studies; we have specifically detailed the step-by-step process to extract data from "Cost-utility studies" and to make it ready to be taken for meta-analysis. Data for the cost-effectiveness of diabetic drug controls are used as a demonstration.

Methods
Basic methods of MA in EESs, including identifying and selecting relevant studies, are similar to other systematic reviews and MAs [5, 17] and should follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [18] guideline when reported. This methodological study was a part of previous MAs, in which some additional specific issues apply to EESs are as follows [19–22]; the relevant protocol was registered in Prospero (PROSPERO 2018 CRD42018105193).

Step 1: Data extraction
All relevant data for comparative EESs (e.g., CUA) should be extracted as follows utilising the Population, Intervention, Comparator, Outcome, and Study type (PICOS) framework:

- General characteristics of EESs including study setting/country, study design (e.g., CUA with model-based, primary CUA alongside RCT/cohort), study perspective, time horizon, discount rate for cost and utility, currency/currency-year, type of EESs [e.g., CUA or cost-effectiveness analysis (CEA)], willingness to pay [WTP; country-specific or gross domestic products (GDP)-base) or country-level cost-effectiveness threshold [23] where appropriate if WTP was not available, and type of economic models.
- Characteristics of patients (P) including indication for treatment, sample size, type of patients (e.g., children/adult, general/specific disease, etc.), mean age, percent male/female, mean body mass index (BMI), etc.
- Type of interventions and comparators (I & C) (with the duration of treatment/dosage/day, etc.)
- Data needed to estimate INB and its variance (O); this includes currency and year, source of cost data (actual data cost collected from patients, central hospital/country costs, etc.), type of cost (e.g., direct/non-direct medical cost, indirect medical cost), and effectiveness outcomes (e.g., life year, QALY, improvement, success, adverse events etc.).

Specific data required for pooling include costs or incremental cost (ΔC), and incremental effectiveness (ΔE) along with their standard deviation (SD), standard

error (SE), or 95% confidence interval (CI) along with covariance between ΔC and ΔE. Some studies may report ICER along with probabilistic sensitivity analysis (PSA). To calculate the INB and its variance, mean and variance of the costs and effectiveness of interventions and comparators along with WTP thresholds are required. In the model-based CUA, studies usually report point estimates of deterministic and/or probabilistic costs and outcomes. We suggest using primarily the measures of central tendency and dispersion measures from PSA results for pooling, as it could better represent a real-life situation considering the distribution of all input variables. Further, to conduct sensitivity analyses using point estimates from the deterministic analysis to see the robustness of results. The WTP threshold was initiated by the Commission on Macroeconomics and Health in 2002 by the World Health Organization CHOosing Interventions that are Cost-Effective (WHO CHOICE) [24]. The WTP threshold in each country usually refers to the standard country guideline based on a fixed value or per capita GDP with returns on investments in health to define whether a health intervention would be (very) cost-effective. [25, 26]. We suggest using the same WTP threshold in monetary units used in the study with further adjustment as per currency conversions as mentioned below. If studies have not reported WTP, one per capita GDP of that study's country and year can be used as WTP along with a sensitivity analysis of three times per capita GDP.

We strongly recommend constructing data extraction forms in advance, a pilot should be performed to make sure that the forms work well and contain all important data specific to that topic.

Step 2: Data harmonisation
Currency conversions
We need to standardize money units usually reported in different currencies (i.e., US $, €, £, ¥) and years by converting to purchasing power parity (PPP) adjusted to US$ for the latest year of analysis[8]. For example, if a study reported cost, ICER, and thresholds in Euros for 2012 and we plan to pool for the current year (e.g., 2022), this currency is firstly converted to 2022 Euros using the historical consumer price index (CPI) of that country (IMF database: https://www.imf.org/en/Publications/WEO/weo-database/2021/October/download-entire-database). Then, the Euro 2022 value to be converted to PPP adjusted US$ rate using conversion rates from the International Monetary Fund [27]. In addition, GDP-based WTP threshold (K) values also need to be corrected for the current CPI 2022 year and PPP; however, standard/country-specific or fixed WTP values only need PPP correction. The variance is calculated as follows:

$$Var_{Adjusted_{2022}} = Var_{Euros_{2012}} x \left(\frac{CPI_{Euros_{2022}}}{CPI_{Euros_{2012}}} x \frac{1}{PPP_{2022}} \right)^2 \tag{1}$$

Estimation of INB and its variance
After currency conversions for cost and K, the INB can be further estimated as follows [8]:

$$INB = Kx\Delta E - \Delta C \tag{2}$$

or

$$INB = \Delta E(K - ICER) \tag{3}$$

Where K is the WTP, and ΔC and ΔE are incremental cost and incremental effectiveness, respectively.

A positive INB favours treatment, i.e., intervention is cost-effective, whereas a negative INB favours the comparator, i.e., intervention is not cost-effective [8, 14, 15].

The variance of INB [8] can be estimated as follows:

$$Var(INB) = K^2\sigma_{\Delta E}^2 + \sigma_{\Delta C}^2 - 2K\sigma_{\Delta E \Delta C} \tag{4}$$

or

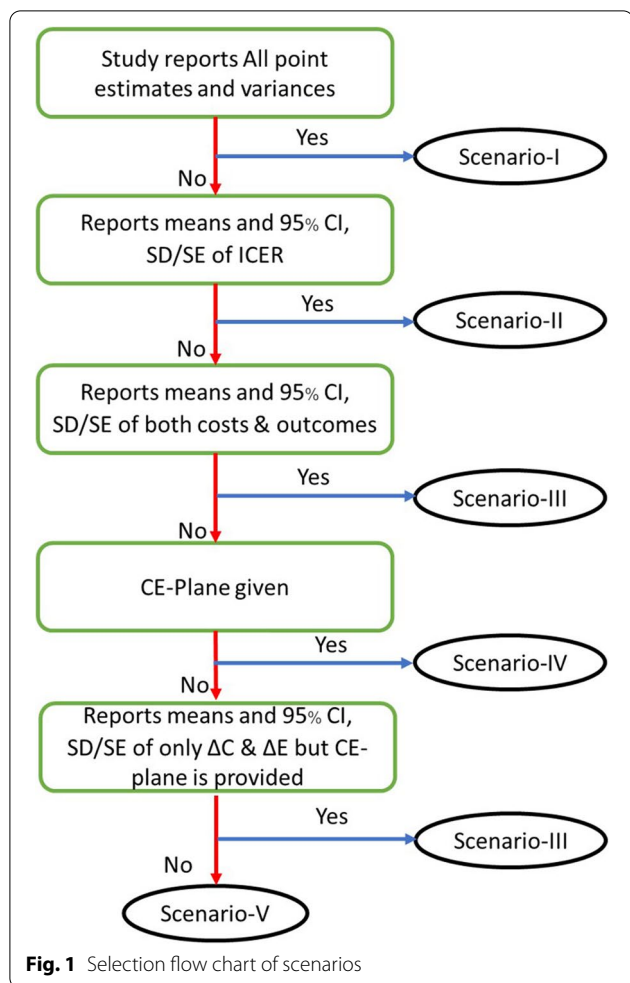$$Var(INB) = K^2\sigma_{\Delta E}^2 + \sigma_{ICER}^2 \tag{5}$$

Where $\sigma_{\Delta C}^2, \sigma_{\Delta E}^2, \sigma_{\Delta E \Delta C}$ are variances of ΔC and ΔE and their covariance, and σ_{ICER}^2 is variance of ICER. However, economic studies can report many different parameters; the five scenarios below, along with a flow in Fig. 1, shows how to obtain INB and variance starting with different reported data [28].

Scenario-1 The primary EES ideally reports the point estimates and variances for every parameter required for the calculation of INB and its variance. The INB can be calculated accordingly to Eqs. (2) to (5).

Scenario-2 The study reports the means and measures of dispersion (95% CIs) of incremental costs & outcomes and ICER. The variance of the ICER can be calculated using the following formulas:

$$95\%CI\,of\,\mu_{ICER} = \widehat{\mu}_{ICER} \pm Z_{a/2}xSE$$
$$UL_{ICER} = \widehat{\mu}_{ICER} + Z_{a/2}xSE$$
$$SE = \frac{UL_{ICER} - \widehat{\mu}_{ICER}}{Z_{a/2}}$$
$$\widehat{\sigma}_{ICER}^2 = SE^2 \tag{6}$$
$$UL_{ICER} = Upper\,limit\,of\,ICER$$
$$Z_{a/2} = Standardize\,normal = 1.96\,if\,alpha = 0.05$$
$$\widehat{\mu}_{ICER} = mean\,ICER$$

Once we know the variance of the ICER, the variance of the INB can be estimated using Eq. (5).

Bagepally *et al. BMC Health Services Research*   (2022) 22:202

Page 4 of 10

Fig. 1 Selection flow chart of scenarios

Scenario-3 The study reports means, and 95% CI, SD/SE of costs/outcomes, or ΔC/ΔE, but does not provide the ICER or its variance. Data for Costs/ΔC and QALYs/ΔE are then used to simulate Costs/ΔC and QALYs/ΔE with 1000 replications using Monte Carlo methods with gamma and normal distributions for Costs/ΔC and QALYs/ΔE, respectively. We suggest simulating costs and QALYs of intervention and comparator for 1000 replications separately; this then leads to calculating the ΔC, ΔE, and covariance between ΔC & ΔE. Data for ΔC and ΔE are assumed to be gamma and normal distributions[29, 30].A sensitivity analysis can be performed using different distributions (e.g., log-normal, exponential for both costs and effective, etc.) to see the robustness of pooling results. The covariance ($\sigma_{\Delta E \Delta C}$) between ΔC and ΔE as well as $\widehat{\sigma}^2_{\Delta C}$ & $\widehat{\sigma}^2_{\Delta E}$ can be then estimated. If the 95% CI is provided, this is converted to SE using Eq. (6) above and used to simulate data. The INB and its variance can be further calculated using Eqs. (2) and (5).

Scenario-4 The study does not report any dispersion but does provide the CE plane graphs, a scatter plot of ΔC on the Y-axis and ΔE on the X-axis, in which individual values of ΔC and ΔE data can be manually extracted from the CE plane using Web-Plot-Digitizer software[31]. Then, means of ΔC, ΔE, and their variances and covariances can be estimated accordingly. Finally, the INB and its variance can be estimated using Eqs. (2) and (5).

Scenario-5 The study reports neither any dispersion nor the CE-plane graph but only provides the deterministic analysis means (or point estimates) of costs, outcomes, and ICER. In such situations, the measures of dispersions can be borrowed from another similar study if they fulfill the following criteria:

- They are in the same stratum of country income,
- Their ICERs are not much different, e.g., ± 50% to 75%
- They are similar in intervention, comparator, time period, country region
- Similar model type and inputs (i.e., discounting, time horizon).

If there is more than one study that meets the criteria, the average of the variances of those studies can be used.

Step 3: Pooling INB
When pooling INBs from many studies, we strongly recommend stratifying by the level of country income, model type, time horizon, and perspective in order to reduce heterogeneity. The country income should be classified as low (LIC), lower-middle (LMIC), upper-middle (UMIC), and high (HIC) as per the World Bank classification[8]. Economic models can include Markov, decision tree, discrete event simulation, or others. Study perspectives should include societal, third-party payer, and patient perspectives. Time horizon should be lifetime (e.g., ≥ 20 or 30 years depending on the disease context) and non-lifetime (e.g., < 5, < 10-years, etc.).

The INB can be pooled across studies using a fixed-effect or a random-effect model depending on the degree of heterogeneity [5, 8, 19, 20, 28].

A) Fixed-effects model

$$\text{INB}_p = \frac{\sum_{i=1}^{S} w_i . \text{INB}_i}{\sum_{i=1}^{S} w_i} \qquad (7)$$

Bagepally *et al. BMC Health Services Research* (2022) 22:202

Page 5 of 10

$$w_i = \frac{1}{Var(INB_i)} \quad (8)$$

B) Random-effects model

$$INB_p = \frac{\sum_{i=1}^{S} w_i^* . INB_i}{\sum_{i=1}^{S} w_i^*} \quad (9)$$

$$w_i^* = \frac{1}{Var(INB_i) + \tau^2} \quad (10)$$

$$\tau^2 = \frac{Q - (S-1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}} \quad (11)$$

where Q is the Cochrane Q test, which has Chi-square distribution; S is a number of included studies/comparisons; τ^2 is a between-study variation.

Similar to MA in other areas, heterogeneity needs to be assessed before pooling INB. Heterogeneity can be visualized by inspection of the forest plot and quantitated using the Cochrane-Q test and the I^2 statistic[5].

$$Q = \sum_{i=1}^{S} w_i \left(INB_i - INB_p \right)^2 \quad (12)$$

$$I^2 = 100\% x \frac{Q - (S-1)}{Q} \quad (13)$$

If heterogeneity is present, i.e., the $I^2 \geq 25\%$ or *p*-value of Q test is less than 0.1; the INBs can be pooled using a random-effects model; otherwise, a fixed-effect model can be applied [8, 19, 20, 28]. Exploring source/s of heterogeneity is strongly recommended. This can be done using a meta-regression to fit each potential source (e.g., time horizon, percent discount rate, threshold values, source of effectiveness measure, risk of bias, economic structure, etc.) one-by-one [8, 19, 20, 28]. If that potential factor explains some proportion of the heterogeneity, including it in the meta-regression model should reduce the I^2 accordingly. There are no established criteria for how much I^2 should be decreased to consider that factor as a significant source of heterogeneity. In our experience, if the I^2 is reduced by about 50% or more from the baseline model (i.e., the model without any factor), such factor/s may be source/s of heterogeneity. A post-hoc subgroup analysis by that factor should be performed accordingly. In addition, sensitivity analyses excluding a few studies with very different characteristics compared to the rest can be used to see if heterogeneity of INBs can be reduced.

Similar to general MA, publication bias should be assessed using a funnel plot and Egger's test. A funnel plot graphs INB estimates on the x-axis against their precision on the y-axis. If all studies are estimating the same true INB, their INBs should be randomly scattered around the true value and form a funnel shape. Egger's test formally tests if the funnel is symmetrical; if this is significant, it usually indicates that there is heterogeneity or missing studies (publication bias) or both. A contour-enhanced funnel plot is further recommended [32]. This plot will contour the area of the funnel into non-significant (*P*-value > 0.05- < 0.1) and significant areas (*P*-value < 0.01 and < 0.05), which will help to differentiate the cause of the asymmetry. For instance, if missing studies fall into the non-significant area, asymmetry might be due to missing studies or publication bias. Conversely, if missing studies are in significant areas, heterogeneity is more likely to be the explanation.

Example

We used data from a MA of CUA of glucagon-like peptide 1 agonists (GLP1) for treatment of type 2 diabetic (T2D) patients who failed to achieve control with metformin monotherapy [19]. A total of 56 studies with 82 comparisons were eligible for pooling INBs. We included comparisons of GLP-1 and dipeptidyl peptidase-4 inhibitors (DPP4i) ($N = 10$); study characteristics are described in Table 1. All studies were from HICs and used country-specific WTP threshold; 9/10 studies used the third-party payer perspective, and 7/10 used a lifetime horizon.

In terms of preparing the data for pooling, 7, 1, and 2 studies provided data matching scenarios 3, 4, and 5, respectively (Table 1). Data for mean cost, QALY, and their incremental values are described in Table 2. Costs and WTP thresholds from each study were converted to $US currency using PPP adjusted for the year 2019 using formula (1).

For the seven studies matching scenario 3 (where mean C and E data along with SDs were reported) (Table 1 and Supl Table 1), the Monte-Carlo method was used to simulate 1000 replicated data based on gamma and normal distributions for cost and QALY data, respectively. Then, ΔC and ΔE along with variance and covariance ($\sigma_{\Delta E \Delta C}$) were calculated. The INB and variance were then calculated following formulas (2) and (4).

The study matching scenario 4 provided CE-plane graphs (Table 1). Data for ΔC and ΔE were directly extracted from the CE plane using Web-Plot-Digitizer [31]. Then, variance and covariance ($\sigma_{\Delta E \Delta C}$) were calculated, leading to estimation of the INB and its variance using formulas (2) and (4).

For the two studies [33, 35] matching scenario 5, the INB variance was adopted from other studies following the steps outlined above. Of the ten included studies, two other studies [36, 40] were conducted in the

Table 1 Selected studies and its analysis scenario

Study	Country	Perspective	Time horizon	Reference year	Threshold	Coversion factor	ICER	Measure of dispersion	CE-Plane	Scenario
Sinha [33]	USA	Payers	Life-time	2008	$ 50,000	1.138489257	yes	none	no	5
Davies [34]	UK	Payers	Life-time	2008	£ 20,000	1.745246374	yes	SD	no	3
Guillermin [35]	USA	Payers	35-yrs	2010	$ 50,000	1.124115573	no	SD	no	5
Lee [36]	USA	Payers	35-yrs	2011	$ 50,000	1.089714997	yes	SD	no	3
Mezquita-Raya [37]	Spain	Payers	Life-time	2012	€ 30,000	1.562673067	yes	SD	yes	3
Steen-Carlsson [38]	Sweden	Societal	Life-time	2013	SEK 500,000	0.116048611	yes	NA	yes	4
Perez [39]	Spain	Payers	Life-time	2012	€ 30,000	1.562673067	yes	SD	no	3
Bruhn [40]	USA	Payers	50-yrs	2014	$ 50,000	1.035412488	yes	SD	yes	3
Roussel [41]	France	Payers	Life-time	2013	€ 30,000	1.261085687	yes	95% CI & SD	yes	3
Barnett [42]	UK	Payers	Life-time	2016	£ 20,000	1.46911077	yes	SD	yes	3

CI confidence interval, *SD* Standard deviation

Table 2 Descriptive of the mean cost and QALY along with their incremental data of comparison between GLP1a vs DPP4i

Author	Cost				QALY			ICER
	Currency	GLP1	DPP4i	ΔC	GLP1	DPP4i	ΔE	
Sinha [33]	US $	170,799	167,163	3636	15.2998	15.3335	-0.0337	-107,893
Davies [34]	£	21,793 ± 544	19,951 ± 521	1842 ± 751	7.52 ± 0.11	7.34 ± 0.11	0.19 ± 0.15	10,158
Guillermin [35]	US $	55,647	57,862	-2215	9.56 ± 0.12	9.28 ± 0.12	0.284 ± 0.172	-7799
Lee [36]	US $	81,444 ± 1079	76,262 ± 1061	5182	8.825 ± 0.117	8.624 ± 0.115	0.201	31,488
Mezquita-Raya[37]	€	54,684 ± 1250	52,387 ± 1346	2297	9.04 ± 0.13	8.87 ± 0.11	0.17	13,266
Steen-Carlsson [38]	SEK	1,360,715	1,304,092	56,624	10.53	10.15	0.38	154,226
Perez [39]	€	56,628 ± 1323	52,450 ± 1394	4177	9.239 ± 0.121	8.838 ± 0.121	0.4	10,436
Bruhn [40]	US $	140,806 ± 1948	138,583 ± 2071	2223	9.618 ± 0.125	9.517 ± 0.130	0.101	22,094
Roussel [41]	€	43,031 ± 1532	40,472 ± 1513	2558 (2427,2689) [a]	10.09 ± 0.13	9.84 ± 0.13	0.25 (0.24, 0.26)[a]	10,275
Barnett [42]	£	24,737 ± 739	22,362 ± 725	2375	9.18 ± 0.12	9.02 ± 0.11	0.15	15,423

Values in cell are mean ± standard deviation, [a]95% CI, *ΔC* incremental cost, *ΔE* incremental QALY, *GLP1a* Glucagon-like peptide 1 agonists, *DPP4i* Dipeptidyl peptidase-4 inhibitors

Table 3 Describe incremental net benefit comparing GLP1i with DPP4i along with variance

Authors	Mean INB (PPP adjusted US $)	Variance INB
Sinha [33]	-6,058	7,58,90,095
Davies [34]	3,063	3,05,70,369
Guillermin [35]	18,452	7,58,90,095
Lee [36]	5,267	7,58,90,095
Mezquita-Raya [37]	1,529	3,66,23,523
Steen-Carlsson [38]	-11,643	4,31,66,49,739
Perez [39]	12,007	7,18,90,710
Bruhn [40]	3,077	9,68,23,864
Roussel [41]	6,373	5,54,03,868
Barnett [42]	1,172	2,45,24,439

INB incremental net benefit, *PPP* purchasing power parity, *GLP1a* Glucagon-like peptide 1 agonists, *DPP4i* Dipeptidyl peptidase-4 inhibitors

USA. For the selection of variance values for the Guillermin et al. study[35], the study period, time-horizon, study perspective, ICER values, drug comparison (Sitagliptin) were most similar to Lee et al. [36] (Table 1 and Table 2). Hence, the INB variance value of the latter was used to estimate the former. The values for Lee et al. also matched the second study [33] most closely.

INB data along with variances are shown in Table 3. The forest plot was constructed by plotting point estimated INBs along with 95% CIs for individual studies (see Fig. 2a); the intervention is cost-effective if the estimated INB falls in the right side of a vertical line of zero and it is not cost-effective if it lines in the left side. These INBs were then pooled across studies using a fixed-effects (inverse variance) model yielding a pooled INB (95% CI) of US$ 4012.21 (-571.43, 8595.85) with I^2 of 0% (see Fig. 2a), this corresponded to individual 95% CIs of INBs which are very much overlapped indicating less likely to be heterogeneous. In the presence of heterogeneity, as indicated by $I^2 \geq 25\%$ or Cochrane-Q $p < 0.1$, a random-effects model (DerSimonian and Laird model) could be used [43]. The pooled INB value is positive but its 95% CI covers 0, i.e., GLP1 agonists might be cost-effective as compared to DPP4 inhibitors but the results did not reach statistical significance.

The robustness of the pooled INB, as well as heterogeneity, can be assessed using various sensitivity and subgroup analyses. Sensitivity analyses omitting the study that used a societal perspective [38] and the study that did not use discounting [40] yielded a pooled INB of US$ 4,032.07 (US$ −554.48, US$8,618.61) and US$4,068.19 (US$ −650.66, US$8787.04), respectively.

The WTP threshold used for these comparisons ranged from US$ 29,382 to US$ 58,024, with a median of US $49,325. Subgroup analyses by median WTP threshold ($<$ vs $\geq$ US $49,325$), time horizon, and source of effectiveness measure were performed (see Table 4), indicating GLP1s were not significantly cost-effective compared with DPP4i in any subgroup. In all these sensitivity and subgroup analyses, the results were similar to the overall pooled INB, indicating that the results are robust.

As in general MA, publication bias was assessed using a funnel plot and Egger's test. There was no evidence of publication bias, seen either by asymmetry on the funnel plot (Fig. 2b) or an Egger's test (coefficient = 0.32, SE = 0.73, $p = 0.672$).
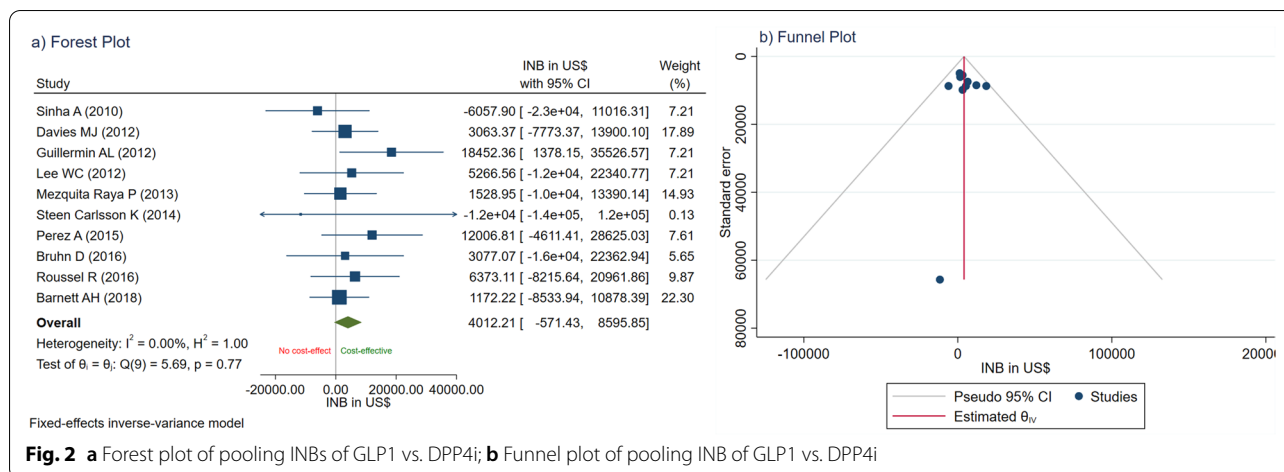
Bagepally *et al. BMC Health Services Research*   (2022) 22:202

Page 8 of 10

Fig. 2 **a** Forest plot of pooling INBs of GLP1 vs. DPP4i; **b** Funnel plot of pooling INB of GLP1 vs. DPP4i

Discussion

We have extended the COMER MA methods for EESs, focusing on data harmonisation; methodological issues include currency, time, discount, perspective, time horizon, and model used to aid in applying a MA for evidence synthesis in EESs. INB and its variance are estimated based on five scenarios. MA is then applied to pool INBs across studies providing a summary estimated CE of treatment relative to control. This evidence should be useful for policymakers in making decisions regarding reimbursement of treatments to a population in countries where resources are limited.

Despite the existence of several guidelines for reporting EESs, studies still vary in how they report the results[44]. This data harmonisation process reported here under the five scenarios can help prepare data to calculate and pool INB values. The different monetary units and year can all be converted to a common standard currency.

We used INB instead of ICER as the economic effect measure following COMER methods because of limitations of the ICER in the ambiguity of interpretation for negative ICER as mentioned above [8, 13]. On the

other hand, positive INBs indicate cost-effectiveness, while negative INBs show non-cost-effectiveness. This information will be required by policymakers [14, 15] in making a decision from both resource-rich and resource-poor countries.

A few challenges should be highlighted when applying a MA in EESs. First, EEs are heterogeneous, which can be caused by model type, population, country income, GDP, perspective, time horizon, and discount rate. We applied the CPI and PPP to harmonise different economic backgrounds as well as the time-lag across the studies [45, 46]. However, it should be noted that using CPI and/or PPP may have some limitations as for the estimation method of price indices, which are calculated from individual prices of only selected commodities rather than all commodities in each country [47]. Considering not only country income but also model type, time horizon, and perspective in stratified analyses may also reduce heterogeneity, if there are sufficient data for stratifying. Furthermore, sub-group and/or sensitivity analyses should be performed to identify specific types of studies/country income where

Table 4 The subgroup analysis results of pooing INB between the GLP1 agonists and DPP4 inhibitors

Subgroup analysis	No. of comparisons	Pooled INB (US$)	95% CI	p-value	I² (%)
Threshold					
< Median ($49,325)	5	3,554.00	-1,825.34 to 8933.34	0.829	0.0
≥ Median ($49,325)	5	5,226.56	-3530.74 to 13,983.86	0.393	2.4
Time Horizon					
Life time	7	2,663.36	-2463.30 to 7790.01	0.852	0.0
Non-lifetime	3	9,386.72	-846.73 to 19,620.17	0.424	0.0
Source of effectiveness					
Multiple study	4	1,538.51	-8,049.60 to 11,126.62	0.742	0.0
Single study	6	4,745.01	-473.58 to 9,963.59	0.534	0.0

Bagepally *et al. BMC Health Services Research* (2022) 22:202

Page 9 of 10

treatments show cost-effectiveness. Economic factors should be considered for subgroups, including WTP, discount rate, type of EES (e.g., within-trial EES versus model-based EES), quality of EESs or risk of bias, the structure of economic model, type of health state, and percent herd immunity for the vaccine, etc.). Different subgroups of these factors may result in different cost-effectiveness findings within HICs and UMICs.

Second, the health EESs are context-specific, usually conducted in individual country settings. However, not all countries have EESs that fit their context because conducting well-designed EESs is very resource-intensive and requires specialised expertise in economic evaluation. Therefore, there will be an even greater need for some systematic synthesis of evidence where resources are limited. Evidence from a MA of EESs will be useful if it is performed with sensitivity to country contexts (e.g., country income, type of model, lifetime, perspective, etc.).

In conclusion, we have described a tutorial of MA in EESs by applying the general methods of MA, additional with specific issues for EESs. The step-by-step approach of data harmonization is demonstrated for facilitating the process of MA. Although evidence of CE is context-specific for each country, conducting such specific individual study is challenging as similar to CE studies due to various practical limitations (e.g., trained manpower, time, resources, etc.). Thus, the MA of EESs should be encouraged; evidence synthesis would be of immense value for the policy decision-making process as well as aid in the comparability of such evidences across countries with similar contexts.

Abbreviations

ΔC: Incremental cost; ΔE: Incremental effectiveness; BMI: Body mass index; CE: Cost effectiveness; CE-plane: Cost effectiveness plane; CI: Confidence interval; CPI: Consumer price index; CEA: Cost effectiveness analysis; COMER: COMparative Efficiency Research; CUA: Cost utility study; DALY: Disability adjusted life year; DPP4i: Dipeptidyl peptidase 4 inhibitors; EES: Economic evaluation studies; GDP: Gross domestic product; GLP1i: Glucagon-like peptide-1; HIC: High income countries; ICER: Incremental cost effectiveness ratio; INB: Incremental net benefit; LIC: Low income countries; MA: Meta analysis; PPP: Purchasing power parity; PRISMA-P: Preferred Reporting Items for Systematic Reviews and Meta-Analyses protocol; PSA: Probabilistic sensitivity analysis; SD: Standard deviation; SE: Standard error; QALY: Quality adjusted life years; UMIC: Upper middle-income countries; US: United states; WHO: World health organisation; WTP: Willingness to pay.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12913-022-07595-1.

Additional file 1.

Author details

[1]Mahidol University Health Technology Assessment (MUHTA) Graduate Program, Mahidol University, Bangkok, Thailand. [2]ICMR-National Institute of Epidemiology, Chennai, India. [3]Social Administrative Pharmacy Division, Department of Pharmacy, Faculty of Pharmacy, Mahidol University, Bangkok, Thailand. [4]Department of Pharmacotherapy, College of Pharmacy, University of Utah, Utah, USA. [5]Centre for Clinical Epidemiology & Biostatistics, School of Medicine and Public Health, Hunter Medical Research Institute, University of Newcastle, New Lambton, NSW, Australia. [6]Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Rama VI Road, Bangkok 10400, Thailand.

References

1. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes, 3. edition. Oxford: Oxford University Press; 2005.
2. Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ. 1986;5(1):1–30.
3. Wisloff T, Hagen G, Hamidi V, Movik E, Klemp M, Olsen JA. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. Pharmacoeconomics. 2014;32(4):367–75.
4. Drummond M, Stoddart G, Labelle R, Cushman R. Health economics: an introduction for clinicians. Ann Intern Med. 1987;107(1):88–92.

Bagepally *et al. BMC Health Services Research*   (2022) 22:202

Page 10 of 10

5. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook.
6. Gomersall JS, Jadotte YT, Xue Y, Lockwood S, Riddle D, Preda A. Conducting systematic reviews of economic evaluations. Int J Evid Based Healthc. 2015;13(3):170–8.
7. Shemilt I, McDaid D, Marsh K, Henderson C, Bertranou E, Mallander J, Drummond M, Mugford M, Vale L. Issues in the incorporation of economic perspectives and evidence into Cochrane reviews. Syst Rev. 2013;2:83.
8. Crespo C, Monleon A, Diaz W, Rios M. Comparative efficiency research (COMER): meta-analysis of cost-effectiveness studies. BMC Med Res Methodol. 2014;14:139.
9. Rehfuess EA, Stratil JM, Scheel IB, Portela A, Norris SL, Baltussen R. The WHO-INTEGRATE evidence to decision framework version 1.0: integrating WHO norms and values and a complexity perspective. BMJ Glob Health. 2019;4(Suppl 1):e000844.
10. Movsisyan A, Rehfuess E, Norris SL. When complexity matters: a step-by-step guide to incorporating a complexity perspective in guideline development for public health and health system interventions. BMC Med Res Methodol. 2020;20(1):245.
11. Shields GE, Elvidge J. Challenges in synthesising cost-effectiveness estimates. Syst Rev. 2020;9(1):289.
12. O'Mahony JF. The Limitations of Icers In Screening Interventions and The Relative Net Benefit Alternative. Value Health. 2015;18(7):A705.
13. Hoch JS, Dewa CS. Advantages of the net benefit regression framework for economic evaluations of interventions in the workplace: a case study of the cost-effectiveness of a collaborative mental health care program for people receiving short-term disability benefits for psychiatric disorders. J Occup Environ Med. 2014;56(4):441–5.
14. Willan AR. Incremental net benefit in the analysis of economic data from clinical trials, with application to the CADET-Hp trial. Eur J Gastroenterol Hepatol. 2004;16(6):543–9.
15. Willan AR, Lin DY. Incremental net benefit in randomized clinical trials. Stat Med. 2001;20(11):1563–74.
16. Miroshnychenko A, Uhlman K, Malone J, Waltho D, Thoma A. Systematic review of reporting quality of economic evaluations in plastic surgery based on the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. J Plast Reconstr Aesthet Surg. 2021;74(10):2458–66.
17. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. Stat Med. 1999;18(3):321–59.
18. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA, Group P-P. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;350:g7647.
19. Bagepally BS, Chaikledkaew U, Gurav YK, Anothaisintawee T, Youngkong S, Chaiyakunapruk N, McEvoy M, Attia J, Thakkinstian A. Glucagon-like peptide 1 agonists for treatment of patients with type 2 diabetes who fail metformin monotherapy: systematic review and meta-analysis of economic evaluation studies. BMJ Open Diabetes Res Care. 2020;8(1):e001020. https://doi.org/10.1136/bmjdrc-2019-001020.
20. Bagepally BS, Gurav YK, Anothaisintawee T, Youngkong S, Chaikledkaew U, Thakkinstian A. Cost Utility of Sodium-Glucose Cotransporter 2 Inhibitors in the Treatment of Metformin Monotherapy Failed Type 2 Diabetes Patients: A Systematic Review and Meta-Analysis. Value Health. 2019;22(12):1458–69.
21. Haider S, Chaikledkaew U, Thavorncharoensap M, Youngkong S, Islam MA, Thakkinstian A. Systematic review and meta-analysis of cost-effectiveness of rotavirus vaccine in low-income and lower-middle-income countries. Open Forum Infect Dis. 2019;6(4):ofz117.
22. IVIR-AC. Meta-analysis of economic evaluations. Wkly Epidemiol Rec. 2021;96(17):138–40. https://www.who.int/publications/i/item/weekly-epidemiological-record-no-17-2021-96-133-144
23. Woods B, Revill P, Sculpher M, Claxton K. Country-Level Cost-Effectiveness Thresholds: Initial Estimates and the Need for Further Research. Value Health. 2016;19(8):929–35. https://doi.org/10.1016/j.jval.2016.02.017.
24. Hutubessy R, Chisholm D, Edejer TT. Generalized cost-effectiveness analysis for national-level priority-setting in the health sector. Cost Eff Resour Alloc. 2003;1(1):8.
25. Robinson LA, Hammitt JK, Chang AY, Resch S. Understanding and improving the one and three times GDP per capita cost-effectiveness thresholds. Health Policy Plan. 2016;32(1):141–5.
26. Marseille E, Larson B, Kazi DS, Kahn JG, Rosen S. Thresholds for the cost–effectiveness of interventions: alternative approaches. Bull World Health Organ. 2015;93(2):118–24.
27. IMF database. Available from: https://www.imf.org/external/pubs/ft/weo/2018/01/weodata/index.aspx
28. Haider S, Chaikledkaew U, Thavorncharoensap M, Youngkong S, Islam MA, Thakkinstian A. Systematic Review and Meta-Analysis of Cost-effectiveness of Rotavirus Vaccine in Low-Income and Lower-Middle-Income Countries. Open Forum Infect Dis. 2019 Mar 8;6(4):ofz117. https://doi.org/10.1093/ofid/ofz117. PMID: 31049363; PMCID: PMC6488528.
29. Polgreen LA, Brooks JM. Estimating incremental costs with skew: a cautionary note. Appl Health Econ Health Policy. 2012;10(5):319–29.
30. Wang H, Zhao H. A study on confidence intervals for incremental cost-effectiveness ratios. Biom J. 2008;50(4):505–14.
31. Rohatgi A. Web Plot Digitizer ver 4.5; 2021. Available from: https://automeris.io/WebPlotDigitizer
32. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol. 2008;61(10):991–6.
33. Sinha A, Rajan M, Hoerger T, Pogach L. Costs and consequences associated with newer medications for glycemic control in type 2 diabetes. Diabetes Care. 2010;33(4):695–700.
34. Davies MJ, Chubb BD, Smith IC, Valentine WJ. Cost-utility analysis of liraglutide compared with sulphonylurea or sitagliptin, all as add-on to metformin monotherapy in Type 2 diabetes mellitus. Diabet Med. 2012;29(3):313–20.
35. Guillermin AL, Lloyd A, Best JH, DeYoung MB, Samyshkin Y, Gaebler JA. Long-term cost-consequence analysis of exenatide once weekly vs sitagliptin or pioglitazone for the treatment of type 2 diabetes patients in the United States. J Med Econ. 2012;15(4):654–63.
36. Lee WC, Samyshkin Y, Langer J, Palmer JL. Long-term clinical and economic outcomes associated with liraglutide versus sitagliptin therapy when added to metformin in the treatment of type 2 diabetes: a CORE Diabetes Model analysis. J Med Econ. 2012;15(Suppl 2):28–37.
37. Mezquita Raya P, Perez A. Ramirez de Arellano A, Briones T, Hunt B, Valentine WJ: Incretin therapy for type 2 diabetes in Spain: a cost-effectiveness analysis of liraglutide versus sitagliptin. Diabetes Ther. 2013;4(2):417–30.
38. Steen Carlsson K, Persson U. Cost-effectiveness of add-on treatments to metformin in a Swedish setting: liraglutide vs sulphonylurea or sitagplitin. J Med Econ. 2014;17(9):658–69.
39. Perez A, Mezquita Raya P. Ramirez de Arellano A, Briones T, Hunt B, Valentine WJ. Cost-Effectiveness Analysis of Incretin Therapy for Type 2 Diabetes in Spain: 1.8 mg Liraglutide Versus Sitagliptin. Diabetes Ther. 2015;6(1):61–74.
40. Bruhn D, Martin AA, Tavares R, Hunt B, Pollock RF. Cost-utility of albiglutide versus insulin lispro, insulin glargine, and sitagliptin for the treatment of type 2 diabetes in the US. J Med Econ. 2016;19(7):672–83.
41. Roussel R, Martinez L, Vandebrouck T, Douik H, Emiel P, Guery M, Hunt B, Valentine WJ. Evaluation of the long-term cost-effectiveness of liraglutide therapy for patients with type 2 diabetes in France. J Med Econ. 2016;19(2):121–34.
42. Barnett AH, Arnoldini S, Hunt B, Subramanian G, Hoxer CS. Switching from sitagliptin to liraglutide to manage patients with type 2 diabetes in the UK: A long-term cost-effectiveness analysis. Diabetes Obes Metab. 2018;20(8):1921–7.
43. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177–88.
44. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, Augustovski F, Briggs AH, Mauskopf J, Loder E, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. BMJ. 2013;346:f1049.
45. Problems and Extensions of PPP, International Finance Theory and Policy - Chapter 30–4: Last Updated on 7/10/99 [http://internationalecon.com/Finance/Fch30/F30-4.php]
46. Burdick C, Fisher L. Social Security cost-of-living adjustments and the Consumer Price Index. Soc Secur Bull. 2007;67(3):73–88.
47. Pigou AC. The Foreign Exchanges. Q J Econ. 1922;37(1):52–74.

Publisher's Note