Amsterdam UMC
Universitair Medische Centra

Cost-effectiveness meta-analysis in systematic reviews of health economic evaluations

Research team

Romain Collet, PhD student

Prof. Dr. Raymond Ostelo

Assoc. Prof. Dr. Hanneke van Dongen

Prof. Dr. Ardine de Wit

Prof. Dr. Judith Bosmans

Prof. Dr. Veerle coupé

Prof. Dr. Ammarin Thakkinstian

Assoc. Prof. Dr. Oraluck Pattanaprateep

Asst. Prof. Dr. Suparee Wisawapipat Boonmanunt

Context - PhD project

TransmUral alLied healthcare Pathway

- Aim:
  - ➢ Address allied healthcare needs
  - ➢ Transition from hospital to home
  - ➢ Multidisciplinary transitional care interventions

Why is TULIP needed?

- Aging population
- Rising prevalence of multimorbidity (38% of adults globally)
- Fragmented care across settings
  - Avoidable readmissions
  - Poorer outcomes
  - Substantial healthcare and societal costs

TULIP planning

- 2023-2024: Needs assessment
  - Systematic reviews
  - Qualitative studies
- 2025-2026: Implementation and evaluation

ELSEVIER

International Journal of Nursing Studies

Volume 168, August 2025, 105103

Cost-effectiveness of multidisciplinary transitional care interventions: A systematic review and meta-analysis

Romain Collet [a d h i 1], Charlotte Groenewoud [i 1], Raymond Ostelo [d h i], Juul van Grootel [b c d], Marike van der Leeden [a c d], Marike van der Schaaf [b c f], Suzanne Wiertsema [a d], Edwin Geleijn [a], Mel Major [e f], Judith Bosmans [g], Johanna van Dongen [d g]

Meta-analysis on cost-effectiveness

- Are multidisciplinary transitional care interventions cost-effective compared to usual care?
- Are the health benefits of MTCI “worth” their additional costs?

Type of studies included

- Economic evaluations conducted alongside RCTs assessing the effectiveness of multidisciplinary transitional care interventions

- Many challenges!

Different challenges

- Developing RQ
- Search strategy (database search, screening, etc.)
- Data extraction
- Quality assessment (risk of bias)
- Data preparation (harmonization of cost-effectiveness estimates from individual studies)
- Meta-analysis
- Generation of cost-effectiveness acceptability curves (CEAC)
- Interpretation (High between study heterogeneity)

Guidance from different papers

ISPOR REPORT · Volume 24, Issue 4, P463-472, April 2021 · ***Open Archive***

Download Full Issue

Critical Appraisal of Systematic Reviews With Costs and Cost-Effectiveness Outcomes: An ISPOR Good Practices Task Force Report

Olena (Lena) Mandrik, PhD [1] · J.L. (Hans) Severens, PhD [2,3] · Ariel Bardach, PhD [4] · ... · Luke Vale, PhD [8] · Torbjørn Wisløff, PhD [9] · Jeremy D. Goldhaber-Fiebert, PhD [10] ... Show more

Guidance from different papers

Research | Open access | Published: 15 February 2022

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

Bhavani Shankara Bagepally, Usa Chaikledkaew, Nathorn Chaiyakunapruk, John Attia & Ammarin Thakkinstian

BMC Health Services Research **22**, Article number: 202 (2022) | Cite this article

Issues with these papers

- ISPOR paper provides general guidance
  - ➢Remains very vague overall
- Bagepally et al. propose creative solutions for harmonizing data
  - ➢But no code is provided so it remains difficult to implement
- Huge gap remains for risk of bias assessment and between study heterogeneity

Development of a tutorial paper

- Step by step guidance for each step
- Clear recommendations for risk of bias assessment and heterogeneity
- Annotated R code for
  - Currency conversion
  - Harmonizing data (scenario’s from Bagepally et al.)
  - Performing meta-analysis
  - Generating CEACs

Quality assessment

- Many people use the CHEERS 2022 statement
  - A set of guidelines for reporting health economic evaluations to ensure they are transparent and useful for decision making
  - Not designed for risk of bias assessment!

- More recently developed: the ECOBIAS checklist
  - Identify and assess the risk of bias in health economic evaluations
  - More suitable

ECOBIAS - advantages

- Comprehensive 22 items checklist addressing various potential biases such as:
  - No treatment comparator bias (no comparator included)
  - Limited scope bias (restricted perspective)
  - Cost omission bias (not all relevant costs included)
  - Selective reporting bias (only favorable results reported)
  - Etc.

- Each item is straightforward to assess

ECOBIAS - issues

- How to provide an overall risk of bias score at the study level?
- How to integrate this score with the interpretation of the pooled results?
- In the literature, researchers use different thresholds (50, 70, 90%)
- Sensitivity analyses are rarely performed. Risk of bias scores are only reported…

Our approach for now

- Comprehensive scoping review of the literature
- ***How do authors of meta-analyses of health economic evaluations assess and incorporate risk of bias and heterogeneity in the interpretation of their findings?***
- Aim: practical recommendations to improve the consistency, transparency, and reproducibility

Outcomes to extract from individual studies

- Costs in intervention and control groups
- Effect: Quality of life OR Quality-Adjusted Life Years (QALYs)
- Measured through self-reported questionnaires pre-post intervention OR extracted from health records

Costs

- Healthcare perspective = direct healthcare costs
- Examples: hospital bills, doctor's visits, surgeries, prescription medications, and diagnostic tests.

- Societal perspective = direct healthcare costs + non-direct
- Productivity loss: Wages lost by patients or caregivers who are unable to work due to illness or injury
- Informal caregiving: The economic value of unpaid care provided by family members and non-professionals.

Effect: QALYs

- Captures both **the quantity of life** (how long a person lives) and the **quality of life** (how good that life is)
- One year lived in perfect health = 1 QALY
- 10 years lived with a quality of life of 0.5 = 10 x 0.5 = 5 QALY

Conversion of QALYs

- Many studies report quality of life (QoL) and not QALYs
- Mapping algorithms exist to convert QoL into a utility score
- Utility score = QoL score ranging from 0 to 1
- Conversion of utility score into QALY assuming a linear increase:

  $QALY = (Utility\_score_{follow\text{-}up} - Utility\ score_{baseline}\ /2) \times Follow\text{-}up\ time$

Considerations

- Many QoL questionnaires exist
- Many algorithms to convert these QoL scores into QALYs
- Difficult to provide standardized guidance in our tutorial
- Source of heterogeneity

Currency conversion

- Conversion to USD $ (most recent year)
- Consumer price indices (CPI): adjusts for inflation within same country
- Purchasing power parities (PPP): adjusts for differences across countries (cost of living)

CPI and PPP conversion formula

$$\text{Cost}_{\text{Target-PPP}} = \text{Cost}_{\text{reported}} \times \frac{\text{CPI}_{\text{origin, ref}}}{\text{CPI}_{\text{origin, reported}}} \times \frac{\text{PPP}_{\text{target, ref}}}{\text{PPP}_{\text{origin, ref}}}$$

- Ref: reference year (e.g., 2025)
- Reported = reported year in the individual study
- Target = target currency (i.e., USD $)

CPI and PPP conversion formula

- For the tutorial, we provide an R code to do this calculation

How to calculate cost-effectiveness?

- INB – Incremental net benefit

- Other option would be the incremental cost-effectiveness ratio (ICER)
  - More difficult to interpret

INB

$INB = (K \times \Delta E) - \Delta C$

K: Willingness to pay threshold (i.e., max amount a health system is willing to pay to increase QALY of 1 unit)

$\Delta C = Cost_{intervention} - Cost_{control}$

$\Delta E = Effect_{intervention} - Effect_{control}$

➢**Positive INB = cost-effective intervention**

Meta-analysis plan

1. Calculate the INB per study separately

2. Pool the different INBs

3. The final result is in USD$ with a confidence interval

Challenge 1

How to calculate the uncertainty around each individual INB?

Some solutions

Research | Open access | Published: 15 February 2022

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

Bhavani Shankara Bagepally, Usa Chaikledkaew, Nathorn Chaiyakunapruk, John Attia & Ammarin Thakkinstian

BMC Health Services Research **22**, Article number: 202 (2022) | Cite this article

INB - Formulas

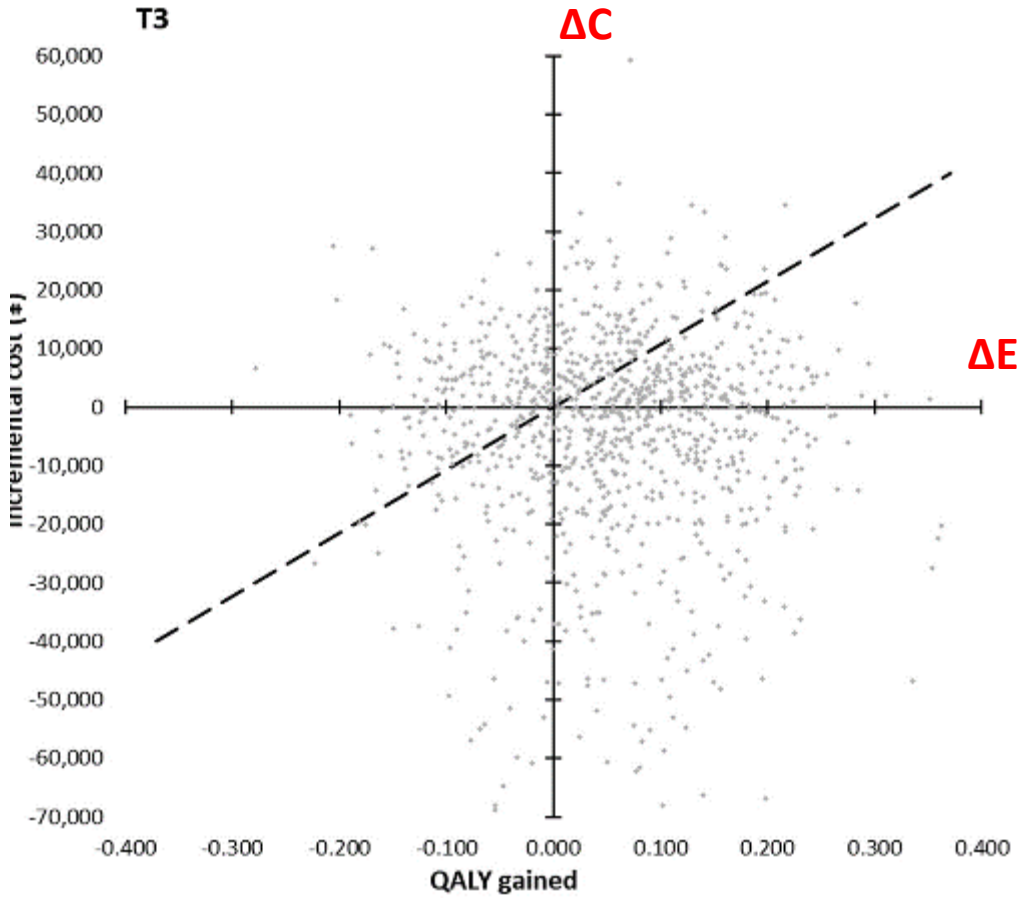
- $INMB = (K \times \Delta E) - \Delta C$

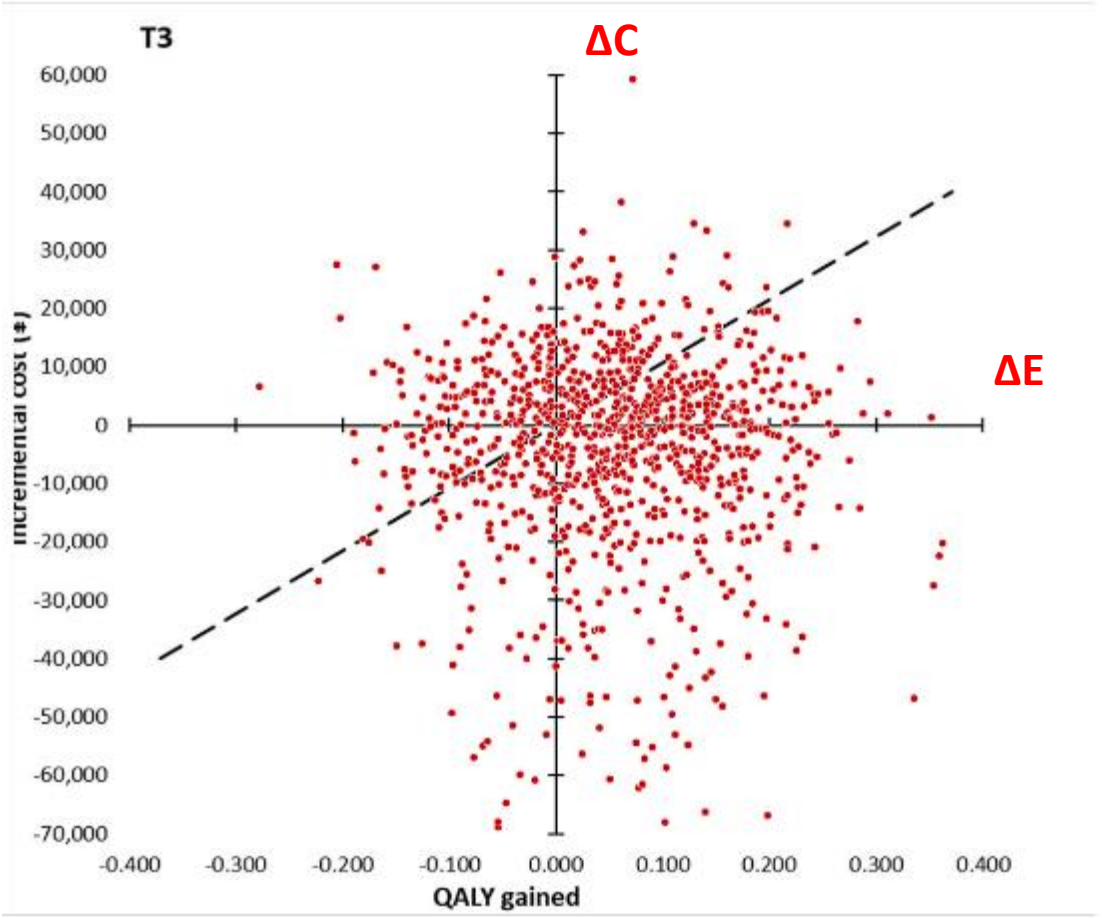
????

- $VAR_{INB} = K^2 \times VAR_{\Delta E} - VAR_{\Delta C} - 2 \times K \times COV_{\Delta E \Delta C}$

Not always reported…

Option 1 – Web Plot Digitizer





Option 1 – Web Plot Digitizer

	ΔE	ΔC
	A	B
1	-0,283359712	6517,786019
2	-0,006968056	6526,722099
3	0,087765331	6538,644358
4	-0,010940232	6552,466663
5	0,033878516	6586,703827
6	0,125673503	6606,697736
7	0,176182706	6646,685554
8	0,054301803	6646,685554
9	0,149006559	6706,667281
10	-0,100114859	6714,109163
11	-0,070484574	6714,464355
12	0,208852409	6726,576399
13	-0,028418456	6728,62129
14	0,096301303	6756,652054
15	-0,07680252	6766,649008
16	0,144895844	6921,607839
17	0,223165929	6927,469271
18	0,132536166	7006,575917
19	0,062216769	7055,803784

Compute (co-)variance

Option 2 – Monte Carlo Simulations

- Simulate ΔC and ΔE with variances reported in articles
- QALYs: Normal distribution
- Costs: Gamma distribution

Option 2 – Monte Carlo Simulations

- Simulate ΔC and ΔE with variances reported in articles
- QALYs: Normal distribution
- Costs: Gamma distribution
- 1000 x simulations = dataset with 1000 x ΔC and ΔE => compute COVARIANCE

Option 2 – Monte Carlo Simulations

- This only works if the correlation between cost and effect is known, which is rarely the case

  - Otherwise, it is also possible to assume that the covariance equals 0

Option 3 – No variances available

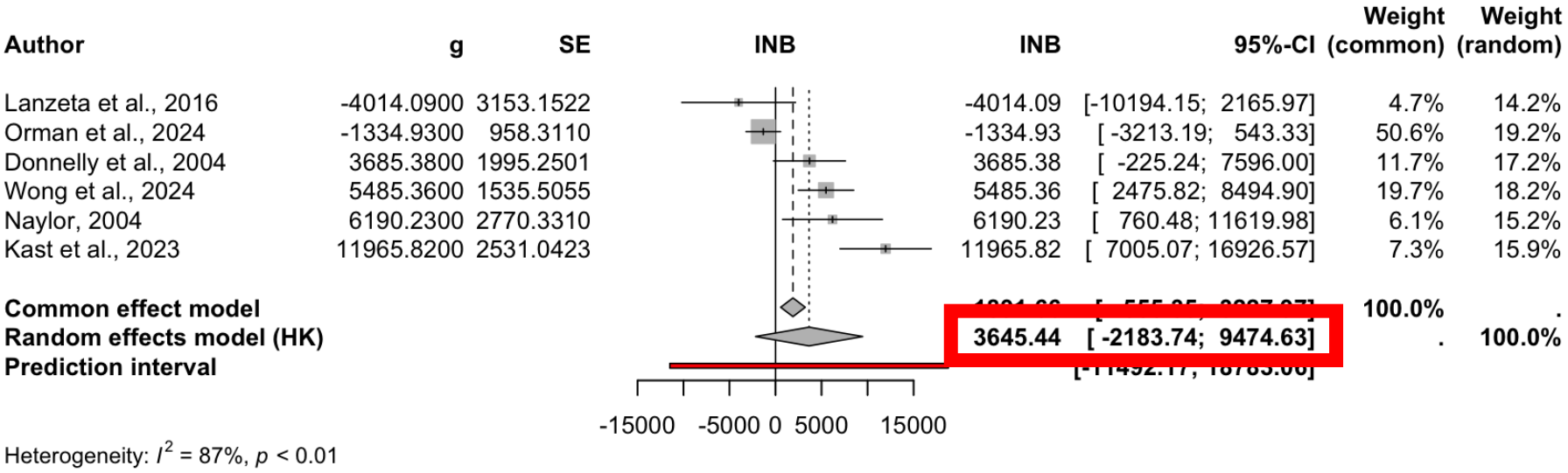
- Imputing the variance of a similar study (similar income level, similar region, similar ICER/INB)

- $Var_{imputed} = [Var_{referent} / mean_{referent}] * mean_{imputed}$

Final dataset

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Author**	**N_exp**	**N_con**	**Exp_costs_M**	**Exp_costs_SD**	**Con_costs_M**	**Con_costs_SD**	**Subgroup_ROB2**	**Subgroup_type**	Incremental_ costs	SD_INC	incremental_ effect	SD_Effect
2	Lanzeta et al., 2016	70	70	8381	7619	6871	7580	moderate	2	1510	10747,35	-0,06	0,7890
3	Orman et al., 2024	251	251	14019,56	33654,79	12005,39	18234,93	moderate	2	2014,16	38277,38	0,01	0,4742
4	Wong et al., 2024	58	58	5212,58	NA	9339,49	NA	moderate	3	-4126,90479	16199,49	0,03	0,0916
5	Donnelly et al., 2004	59	54	16006,21	13038,12	19402,57	12031,11	high	3	-3396,35921	17740,92	0,01	0,2567
6	Kast et al., 2023	109	119	43377	11148	54890	13793	moderate	2	-11513	17740,92	0,01	0,0364
7	Naylor et al., 2004	118	112	8977,44	NA	14673,57	NA	moderate	2	-5696,13	34908,61	0,015	0,0364

Meta-analysis



- In the tutorial article, we will provide the R code to generate such forest plot

Challenge 2

But what about **K**?

Remember

INB = (K x ΔE) - ΔC

K: Willingness to pay threshold (i.e., max amount a health system is willing to pay to increased QALY of 1 unit)

Each country has its own K threshold! Big source of heterogeneity

Bagepally et al.

- Country-specific K for each INB
- Maybe not the best option…

Research | Open access | Published: 15 February 2022

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

Bhavani Shankara Bagepally, Usa Chaikledkaew, Nathorn Chaiyakunapruk, John Attia & Ammarin Thakkinstian

BMC Health Services Research **22**, Article number: 202 (2022) | Cite this article

Solution

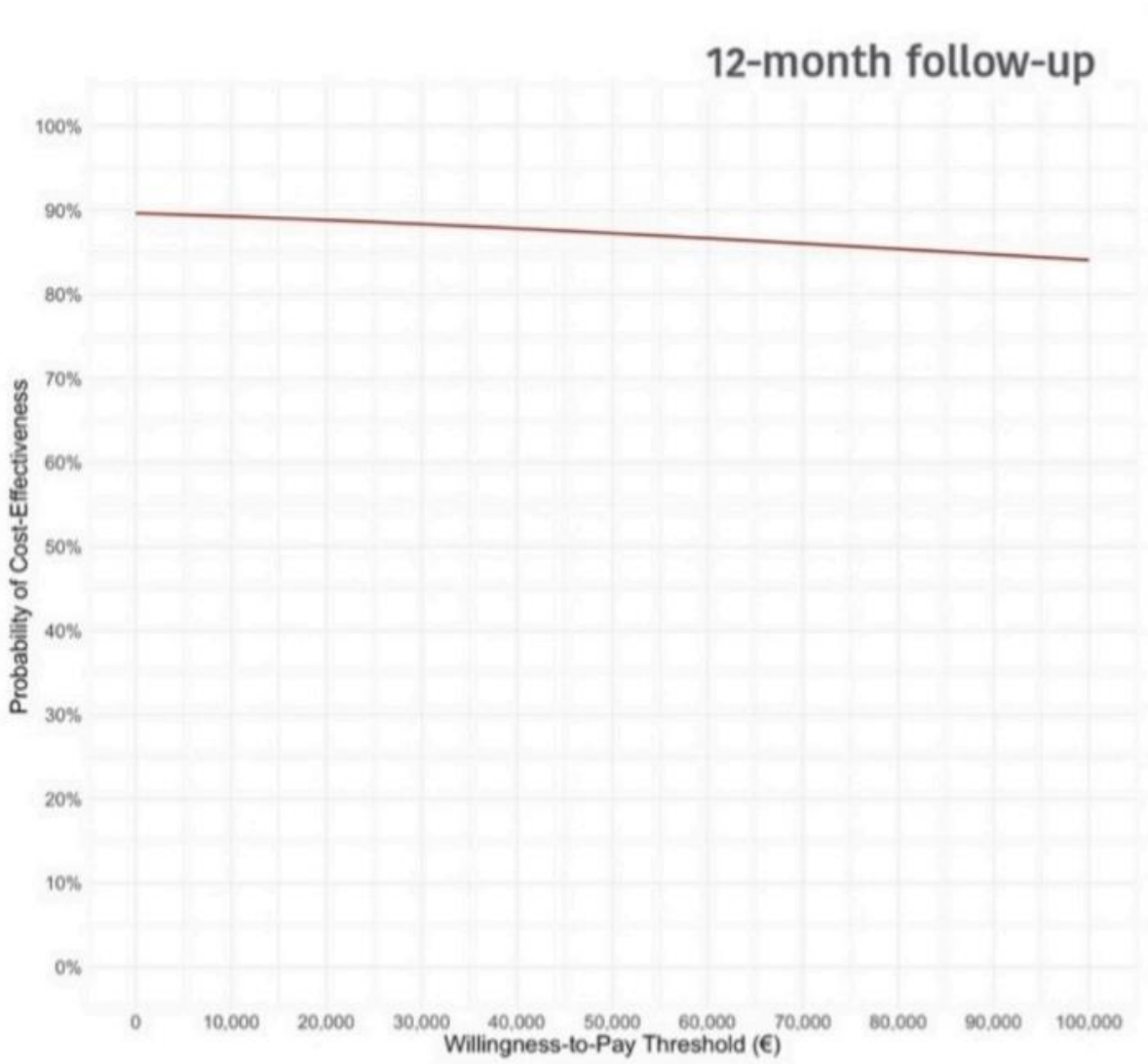
- Perform many meta-analyses with different values of K
- Plot the results on a graph where policymakers from different countries can see what the probability of cost-effectiveness is for their situation
- 21 meta-analyses with K= $ 5000, $ 10000, $ 15000, …, $100000

Perspective	Follow-up duration	# studies	N	K threshold (€)	Outcome (95 % CI)	Probability of cost-effectiveness	Heterogeneity (I^2)
Incremental net monetary benefit (INMB)							
Healthcare	12 months	6	1329	0	3452 (−1912; 8816)	90 %	100 %
				5000	3467 (−1964; 8899)	89 %	100 %
				10,000	3482 (−2020; 8985)	89 %	100 %
				15,000	3497 (−2079; 9074)	89 %	100 %
				20,000	3512 (−2143; 9166)	89 %	99 %
				25,000	3526 (−2209; 9261)	89 %	99 %
				30,000	3540 (−2279; 9360)	88 %	99 %
				35,000	3554 (−2353; 9461)	88 %	99 %
				40,000	3567 (−2430; 9565)	88 %	99 %
				45,000	3581 (−2510; 9672)	88 %	98 %
				50,000	3594 (−2594; 9782)	87 %	98 %
				55,000	3607 (−2681; 9895)	87 %	98 %
				60,000	3619 (−2771; 10,010)	87 %	98 %
				65,000	3632 (−2864; 10,128)	86 %	98 %
				70,000	3644 (−2960; 10,249)	86 %	98 %
				75,000	3656 (−3059; 10,372)	86 %	97 %
				80,000	3669 (−3160; 10,497)	85 %	97 %
				85,000	3680 (−3264; 10,625)	85 %	97 %
				90,000	3692 (−3370; 10,755)	85 %	97 %
				95,000	3704 (−3479; 10,887)	84 %	97 %
				100,000	3715 (−3590; 11,021)	84 %	97 %

Plot CEAC curves

- Plot the results on a graph where policymakers from different countries can see what the probability of cost-effectiveness is for their situation

- In the tutorial, we will provide the R code to generate these curves


12-month follow-up
Probability of Cost-Effectiveness
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0
10,000
20,000
30,000
40,000
50,000
60,000
70,000
80,000
90,000
100,000
Willingness-to-Pay Threshold (€)

Challenge 3

How do we interpret these results?

Interpretation of pooled results

- Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach
- Certainty of evidence **rather than solely focusing on statistical significance**
- 5 criteria
- Certainty level ranges from “very low” to “high”

GRADE

1. Overall risk of bias (more than 50% high-risk-of-bias studies?)
2. Consistency (of the effects across studies, evaluated with the I^2 statistic)
3. Precision (width of confidence intervals)
4. Directness (evidence obtained from different populations)
5. Publication bias (funnel plots, Egger’s test)

GRADE assessment - challenges

- Overall risk of bias: How?
- Consistency: Usually very high heterogeneity…
- Precision: wide confidence intervals because of skewed costs…
- **There is no clear GRADE guidance for cost-effectiveness meta-analysis!**

In our previous meta-analysis

- Overall risk of bias: We used the RoB 2 tool combined with CHEERS
- Consistency: We systematically downgraded the certainty level
- Precision: instead of looking at the confidence intervals we looked into cost-effectiveness-probabilities and downgraded when these were <80%

Sources of heterogeneity - stratification

- Patient populations
- Country income level
- Variability of the interventions (when not a drug intervention)
- Different contexts (healthcare systems)

Next steps

- Scoping review on risk of bias assessment and heterogeneity
- Complete R scripts for each step
- Deadline for the tutorial end of March 2026

Usability and implementation of methodological tutorials

A three-step test interview study

Research question

- ***What is the practical value of methodological tutorials for HTA and applied clinical researchers and what are facilitators and barriers to their implementation in day-to-day research practice?***

Participants

- Researchers (PhD students, assistant professors, associate professors, professors) with an interest in health economic evaluation

- Not expert in statistical programming (in R)

Procedure

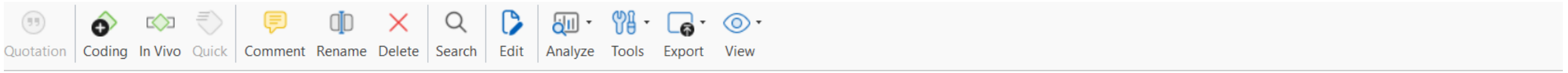
- We provided an example tutorial to participants (an R tutorial on trial-based economic evaluations)
- Participants performed the tutorial in front of us, while thinking out loud
- We then conducted a semi-structured interview on their experiences with the tutorial, and their recommendations for improving future tutorials

Additional semi-structured interviews

- Authors that had cited the tutorial we used for the experiment
- Insight into how they adapted the tutorial to their own data

Thematic analysis

1. Familiarization with the data
2. Initial coding
3. Theme identification
4. Review and refinement of themes
5. Defining and naming themes
6. Integration of themes into a coherent narrative



that you've used the paper that they, the tutorial that they wrote on trial-based economic evaluations. And so my first question is how did you use this tutorial? Did you use all of it? Did you use the codes, or some parts of it?

2

Yeah, I have to go back in time a little bit because the first time I used it is already. Yeah, I don't know, when it just came out. I think two years ago, 1 1/2 years ago maybe. So I did the whole. I tried to do their course, but, their workshop during a conference. But I was, I think suddenly unable to attend, so I never did the full course, but I did receive the materials. Before, I think, the official documents came out, so I'm just thinking which version I used. But anyway what I did was I re-read their whole paper. Then I also did all the test file like the GitHub and yeah, I just followed the steps to see if I could do it on their test file. So I practised first without any of my own data.

And after that what I did was basically using their syntax, so their codes and then as a starter, and then adjust them for my project.

So and now what I do is I keep sort of opening the most recent project probably and then yeah, adjusting it because I of course often have much, many more measurements and more variables and more extensive datasets than their example.
But I still have on top that it's based on their tutorial and then I give it my project name, but I yeah [stuttering] the basis for my codes. Is their codes yeah.

1

Yeah. OK. So then you would say that this was a good basis for starting and yeah and?

2

9:1 Th... The exercise part can be ...g an economic evaluation

9:... Adapting the code to own context/ needs- more informati

9:3 but... The tutorial contains the b...valuation and this is fine!

Theme identification

- Excel

Results - participants

- 20 participants
- 17 in the Netherlands, and 3 in Thailand
- The majority reported being beginners or unfamiliar with R (82%) and economic evaluations (88%)

Results - qualitative synthesis

- Theme 1: Getting in and getting it to run
- Theme 2: Understanding and making sense of what tutorials teach
- Theme 3: Adapting and applying tutorials in practice

Theme 1: Getting in and getting it to run

- Beginners struggled with installation, file access, and navigating the software environment

- Managing errors and uncertainty during code execution requires informal support from colleagues, online searches, or generative AI tools

Example quotes

“I don’t know how to interpret these conflicts that appear in the console. I’m not sure whether everything is correct or whether something needs updating. In a normal situation, I would just ask ChatGPT what it means.” (Participant #5)

“What I liked was that the tutorial first explained the theory and then moved to the practical part. I could link the two very easily, which was really helpful. The R code itself was also straightforward…clicking the green button and running it was not difficult.” (Participant #4)

Theme 2: Understanding and making sense of what tutorials teach

- Clear language, well-commented code, and justification of methodological choices increased confidence
- Matching their outputs to the manuscript figures
- Trust in the authors of the tutorial

Example quotes

“For concepts like utility values or QALYs […] It would help to include a small box explaining the most common terms, so I understand them in your context and not just whatever definition Google gives me […] Explaining the ten most important concepts would be very valuable.” (Participant #1)

“It was very easy to run the analysis, but to interpret the results…[I would need additional information]” (Participant #3)

Theme 3: Adapting and applying tutorials

- Adapting the code to own data = substantially more difficult than completing the tutorial itself
- Hesitant to modify scripts without guidance
- Despite these challenges, participants valued tutorials as efficient learning tools and useful starting points for applied research

Example quotes

“I would recommend it [this tutorial] to people who have good background knowledge, so they understand what happens in each step and can integrate the statistics […]. They don’t necessarily need a health economics background […] a background in epidemiology is fine, I think.” (Participant #14)

“After reading this article and completing the tutorial, I would still need to study more the theory behind. If I had to run a trial-based economic evaluation […] I would find this [the tutorial] super useful […] but at my level I still need a better grasp of the concepts because I am not fully understanding what I’m doing in R.” (Participant #5)

I am curious…

- What do you think of methodological tutorials?
- Who do you think should use – or should not – use them?
- Do you have any recommendations for making them more accessible?