PharmacoEconomics
https://doi.org/10.1007/s40273-026-01628-x

SYSTEMATIC REVIEW

Applications of Structural Expert Elicitations for Economic Evaluations: A Systematic Review Update

Benjamin P. Geisler[1] · **Fredrik Holmboe**[2] · **Irene Starinieri**[1,3] · **Eline Aas**[1,4]

Received: 21 December 2025 / Accepted: 14 May 2026

Abstract

Background Structured expert elicitation (SEE) has become increasingly important in health technology assessment and economic evaluations. Complementing previous work, we aimed to synthesize recent developments in published SEE applications within health economics over the past 8 years.

Methods A systematic literature search was conducted in Medline and Embase databases from April 2017 to February 2026, supplemented with snowball sampling, to identify applications of SEE as part of economic evaluations. Data extraction and synthesis focused on expert selection, elicitation methods, and analytical techniques to identify commonalities and gaps.

Results In total, 28 studies met the inclusion criteria. SEE applications covered diverse health interventions, from rare diseases treatments to diagnostic accuracy assessments. The number of experts recruited through purposive sampling varied from 1 to 18 clinicians per study. SEE processes remain bespoke and diverse, spanning from paper-based to software-assisted remote techniques. The studies used mainly variable and fixed interval methods (29% versus 67%) for encoding. Aggregation methods were mainly mathematical, with some studies using consensus approaches. Most studies (75%) directly incorporated pooled expert distributions into decision models.

Conclusions While SEE methods vary considerably across applications, suggesting that optimal approaches have yet to emerge, there is growing recognition of their potential for informing healthcare decision-making where empirical data are scarce, particularly in rare diseases and early-stage technology assessment. Future research should prioritize standardizing best practices, validating expert predictions against subsequently available empirical data, and developing enhanced bias mitigation strategies to improve the credibility of expert-informed health economic evaluations.

✉ Benjamin P. Geisler
b.p.geisler@medisin.uio.no

[1] Department of Health Management and Health Economics, University in Oslo, Postboks 1089 Blindern, 0317 Oslo, Norway

[2] AbbVie AS, Lysaker, Norway

[3] Erasmus University, Rotterdam, Netherlands

[4] Norwegian Institute of Public Health, Oslo, Norway

Published online: 19 June 2026

Key Points for Decision Makers

When robust clinical trial data are unavailable, e.g., for rare diseases or small patient populations, structured expert elicitation offers a transparent and increasingly accepted method for filling evidence gaps in health economic models. Decision-makers can have greater confidence in reimbursement and resource allocation decisions supported by this approach when it is applied systematically.

Despite growing use, the way expert knowledge is collected and combined still varies considerably across studies. Payers and health technology assessment bodies should carefully scrutinize how expert input was gathered and processed, as methodological choices can meaningfully influence model outputs and cost-effectiveness estimates.

To strengthen future evaluations, there is a clear need for more consistent standards governing how expert elicitation is conducted and reported. Establishing such standards would improve the comparability and credibility of submissions, ultimately supporting more informed and defensible coverage and pricing decisions.

1 Introduction

Incorporating information from experts has been part of model-based economic evaluation for a long time [1]. In healthcare, clinicians have traditionally played a central role in validating and verifying clinical components of models—such as treatment pathways and treatment duration and dosage—in support of decision-making. Historically, this involvement has been largely pragmatic and informal, with limited use of structured frameworks for eliciting and reporting expert input [2–4]. Over the last two decades, however, this has changed, and expert elicitation has become more formalized and methodologically robust [2, 4–7]. The need for expert information is also increasing, partly due to the growing number of health technologies targeting smaller populations, for which trial-based empirical data are often limited or unavailable.

Structured expert elicitation (SEE) can be understood as a framework for a systematic approach for eliciting information from experts following recommendations and guidelines [5, 8]. Over the past 5–10 years, there has been a growing emphasis on articulating and standardizing best practices for SEE, as reflected in the publication of guidelines such as the 2023/2024 Professional Society for Health Economics and Outcomes Research (ISPOR) report and others [5, 8–11]. Existing recommendations and guidelines vary in scope and level of detail, but the most comprehensive documents typically cover the number of experts to include, recommended qualifications and experience, the content and format of the information dossier provided to experts, tools and procedures for collecting expert elicitation (both individually and in consensus settings), and methods for aggregating and extrapolating elicited data.

Over the same period, there has also been notable growth in SEE frameworks, tools, methods, and protocols used in health economics. In addition to longstanding approaches such as the modified Delphi technique, more recently refined or developed elicitation frameworks include the Sheffield Elicitation Framework (SHELF), Cooke's classical method, the Investigate–Discuss–Estimate–Aggregate (IDEA) protocol, the Medical Research Council (MRC) guidance, and the Expert Elicitation Tool (EXPLICIT) reference protocol [8, 12–15]. Furthermore, recent developments include the emergence of web-based elicitation platforms that facilitate remote expert engagement and the increased integration of SEE into Bayesian analytical frameworks.

SEE is particularly valuable for healthcare decision-makers when evidence is sparse or uncertain, for example, in rare diseases, small patient subgroups, emerging therapeutic areas, and early-stage health technology assessments where timely evidence generation through traditional empirical methods may be challenging or impossible. As a result, health technology assessment (HTA) agencies have increasingly recognized the role of SEE and begun to incorporate it into their guidelines. Agencies such as the National Institute for Health and Care Excellence (NICE), the Canadian Agency for Drugs and Technologies in Health, the Pharmaceutical Benefits Advisory Committee, Haute Autorité de Santé, Zorginstituut Nederland, Autoridade Nacional do Medicamento e Produtos de Saúde, and the Scottish Medicines Consortium have all integrated SEE into their guidelines or methodological frameworks, while others are considering or preparing to adopt SEE methods [5, 16].

To date, four systematic reviews have examined the use of SEE methods in decision-analytic modeling for HTA and economic evaluation [17–20]. However, these reviews have limitations. The most recent review focused on immature survival data [20]. Another recent review conducted systematic searches in MEDLINE and Web of Science but did not include EMBASE and might therefore have missed eligible studies [19]. The two earlier reviews only considered studies published up to 2017 [17, 18]. Consequently, despite individual studies describing specific applications or methodological innovations, there is currently no review that identifies and synthesizes applications of SEE methods for HTA from 2017 onward. A comprehensive summary

and up-to-date overview would be valuable for identifying emerging best practices, highlighting priorities for further methodological research, and supporting the consistent and appropriate use of SEE across diverse healthcare decision-making contexts.

Therefore, the aim of this study was to systematically identify and review applications of SEE in model-based economic evaluations of healthcare interventions published in peer-reviewed journals since 2017. We were particularly interested in whether, and in what ways, recent applications of SEE differ in their methodological approaches from patterns reported in earlier reviews.

2 Methods

We updated the systematic literature search conducted by Soares et al. in 2017 and published in 2018 [18]. We accessed both the Medline and Embase databases for entries from April 2017 until the day of our search, 5 November 2024. An updated search was run on 3 February 2026. The search strategy encompassed a combination of Medical Subject Headings (MeSH), Emtree terms, and other pertinent keywords via Ovid's proprietary mapping technology. In addition, we identified papers through "snowballing" (i.e., references cited within identified studies). These articles were reviewed and, if relevant, included in the review. See supplementary materials for the search syntax, the database fields searched, and other details.

Two reviewers (B.P.G., I.S.) reviewed titles and abstracts as well as full texts independently; disagreements were resolved through discussion. To ensure comparability with previous systematic reviews and so that results tables can simply be combined, we used the same selection criteria and the same three extraction tables as Soares et al. 2018 [18]. We included studies that were original research and that were published as a full-text manuscript describing in sufficient detail an application of an SEE process (or referencing such a process) used to inform model-based cost-effectiveness and/or cost–utility analyses of healthcare interventions either independently or as part of an HTA. In addition to Soares et al., see also supplementary information for the operationalized in- and exclusion criteria. Table 1 provides an overview of the studies fulfilling the inclusion criteria, including information on the type of health intervention under investigation, whether it involved an early HTA assessment (versus a reimbursement decision), as well as the types of input parameters elicited; Table 2 lists the professional background and expertise of consulted experts, the encoding method (i.e., variable interval methods [VIM] and fixed interval methods [FIM]), detailed information on how experts provided their input, the method of aggregation (mathematical or behavioral), and the weighting strategies including whether equal weights, performance-based, or objective weighting approaches are reported; and Table 3 includes details on how the SEE process was conducted, opportunities for revisions, what technical tools or software were used, training and piloting, the analytic technique used—e.g., pooling or fitting—and if and how the pooled distribution was implemented into the decision-analytic model. The classification of studies as supporting "early assessment" versus "reimbursement" decisions is based on how studies characterized their own objectives within peer-reviewed publications and may not reflect the full landscape of SEE use across the technology development and assessment lifecycle. Based on the development over the last years, we replaced the category "format/software" in Soares et al. with specific SEE framework, including MRC, the modified Delphi method, Cook's classical methods, IDEA, and Expert Elicitation Tool (EXPLICIT) [12, 14, 15, 21, 22]. Search results were processed and managed in Covidence (Level 9, Melbourne, Victoria, Australia).

3 Results

We initially identified 186 studies through the systematic literature searches. After deduplication and title and abstract screening, 75 potentially relevant studies underwent review of their full texts, resulting in 18 remaining studies that matched the inclusion criteria. Through snowballing, we identified an additional four studies. In a search update in early 2026, we identified another seven studies, for a total of $n = 28$ identified studies [15, 23–49]. See Fig. 1 for a PRISMA flow chart for the consolidated numbers.

The herein identified studies cover a diverse array of health care interventions, with treatments representing the generally most common subject of the economic evaluation (19 studies; 70%), followed by diagnostic or screening interventions (seven studies; 26%), and one study focusing on methodological bias adjustment [39]. Notably, several studies concerned rare diseases or specialized populations, such as pediatric hemophilia [37], DiGeorge syndrome [26], and advanced cutaneous squamous cell carcinoma [31], using SEE to address the paucity of empirical data in these conditions. The parameters targeted for elicitation varied considerably across studies, though event and transition probabilities were the most frequently elicited ones, appearing in approximately half of the studies. Clinical effectiveness parameters, including sensitivity and specificity for diagnostic tests, survival probabilities, and progression-free survival rates, constituted another major category. Finally, resource use, costs, and utility values were elicited less frequently, appearing in only five studies. Half (14 studies: 50%) of the included studies were conducted to inform early assessments during research and development phases rather than

Table 1 Health interventions, timing of the assessment, and parameter types

Study	Type of strategy under investigation	Was the aim to inform an early assessment (i.e., R&D) rather than reimbursement?	Types of parameter(s) elicited
Cox et al. (2026) [43]	Treatment; TMS, a therapeutic neuromodulation intervention for treatment-resistant depression (TRD)	Early assessment	Clinical effect parameters, disease progression dynamics (time profile of improvement under TAU), resource use
Lee et al. (2025) [46]	Treatment, systematic therapies for advanced metastatic RCC	Reimbursement	Long-term survival; conditional PFS; survival proportions
Lim et al. (2025) [47]	Maintaining an active hospital microbiology laboratory service (blood culture + AST)	Early assessment	Event/transition probabilities including probabilities of clinical condition (improvement, deterioration, remaining unchanged) and probabilities of antibiotic prescribing decisions
Wang et al. (2025) [48]	Treatment/service uptake strategy: strategy concerns exercise interventions in women with early-stage endometrial cancer	Early assessment	Relative treatment effectiveness (relative risk reduction of cardiovascular events due to exercise) and diffusion/implementation parameters (clinician referral probability, patient uptake probability)
Westwood et al. (2024) [23]	Accuracy and cost-effectiveness of stroke imaging	Reimbursement	Sensitivity and specificity for large-vessel occlusions
Borsoi et al. (2023) [28]	Osteoporosis diagnosis via REMS ultrasound	Reimbursement	Healthcare resource consumption
Di Fusco et al. (2023) [27]	COVID-19 vaccines	Early assessment	Relative importance of broader value elements (e.g., broader vaccine value elements), prioritization, challenges/recommendations for economic evaluation
Gorelova et al. (2023) [26]	Artificial thymus for DiGeorge syndrome	Early assessment	Event probabilities (e.g., eligibility, success, mortality, and survival rates)
Gray et al. (2023) [25]	Prevention of hospital-acquired hypoglycemia	Reimbursement	Event probabilities (hypoglycemia prevention by patient complexity)
Squires et al. (2023) [24]	Diagnostic/screening in the ED; time and resource allocation for low-acuity attenders	Reimbursement	Time durations of ED processes
Willigers et al. (2023) [49]	Survival extrapolation strategy; strategy concerns long-term survival modelling beyond trial horizon using Bayesian methods informed by formal expert elicitation in CKD	Early assessment	Long-term survival probabilities
Ayers et al. (2022) [29]	CAR-T therapy for relapsed/refractory MM, a rare disease	Early assessment	OS
Paly et al. (2022) [44]	Treatment/service uptake strategy; strategy concerns survival extrapolation methods for immune checkpoint inhibitor therapies in advanced melanoma	Early assessment	Survival probabilities at 10- and 20-year landmarks and survival times
Klijn et al. (2021) [45]	Survival extrapolation strategy; comparison and retrospective validation of alternative survival modelling approaches for immuno-oncology therapies using trial data	Reimbursement	Long-term survival expectations used for validation (landmark survival at 10 and 20 years, mean survival estimates)
Konidaris et al. (2021) [31]	Treatment; advanced cutaneous squamous cell carcinoma	Reimbursement	PFS, OS, AEs, utility decrements, costs
Petersohn et al. (2021) [30]	Treatment; pharmacological treatment for PAD	Reimbursement	Clinical effectiveness, AEs, QoL impacts, costs

Table 1 (continued)

Study	Type of strategy under investigation	Was the aim to inform an early assessment (i.e., R&D) rather than reimbursement?	Types of parameter(s) elicited
Sankatsing et al. (2020) [40]	Digital breast tomosynthesis versus digital mammography for breast cancer screening	Early assessment	Sensitivity, PPV of recall and screening costs with digital breast tomosynthesis
Stevenson et al. (2020) [33]	Diagnostic accuracy in neurodegenerative diseases such as CJD	Reimbursement	Event probabilities, specifically probability of misdiagnosis by age
Vokó et al. (2020) [32]	Treatment of negative symptom schizophrenia	Reimbursement	Transition probabilities in a DA model for schizophrenia health states
Cope et al. (2019) [36]	Treatment of CAR-T for pediatric B-cell ALL	Reimbursement	2-, 3-, 4-, and 5-year survival probabilities
Grand et al. (2019) [41]	Treatment of endometriosis-related pain	Early assessment	Transition probabilities, utilities, and costs
Hitimana et al. (2019) [35]	WHO antenatal care recommendations	Reimbursement	Event probabilities (e.g., estimated reductions in perinatal and maternal mortality), relative effectiveness
Rossi et al. (2019) [34]	Screening and diagnostic pathway for asymptomatic RCC via ultrasound	Early assessment	Stage distribution of undiagnosed RCC cases, non-attenders, FNR, diagnosis probability by stage
van Hoeven et al. (2019) [42]	Anti-B19V-tested blood products for transfusion	Early assessment	Proportion of complications, morbidity/mortality
Kip et al. (2018) [38]	Rapid exclusion of NSTEMI using a triple-biomarker strategy	Early assessment	Sensitivity, specificity, NPV, and impact on discharge rates and interventions
Rodriguez et al. (2018) [37]	Treatment of pediatric hemophilia joint bleeding with first-line agents (rFVIIa or pd-aPCC)	Reimbursement	Treatment dose and frequency, bleed resolution rates at 24/48 h, rebleeding rates
Grigore et al. (2017) [15]	PCPs' advice on alcoholic steatosis	Early assessment	Patient response rates
Schnell-Inderst et al. (2017) [39]	Treatment effect bias adjustment for meta-analyses of RCTs and observational studies	Reimbursement	Bias-adjusted treatment effect estimates (RRs, 95% CIs) for both internal and external biases

AE adverse event, *ALL* acute lymphoblastic leukemia, *AST* antimicrobial susceptibility testing, *B19V* parvovirus B19, *CAR-T* chimeric antigen receptor T-cell (therapy), *CI* confidence interval, *CJD* Creutzfeldt–Jacob disease, *CKD* chronic kidney disease, *COVID-19* coronavirus disease 2019, *DA* decision-analytic, *ED* emergency department, *FNR* false-negative rate, *MM* multiple myeloma, *NSTEMI* non-ST-elevation myocardial infarction, *NPV* negative predictive value, *OS* overall survival, *PAD* peripheral artery disease, *PCP* primary care physician, *PFS* progression-free survival, *PPV* positive predictive value, *QoL* (health-related) quality of life, *RCC* renal cell carcinoma, *RCT* randomized controlled trial, *R&D* research and development, *REMS* radiofrequency echographic multispectrometry, *RR* risk ratio, *TAU* treatment as usual, *TMS* transcranial magnetic stimulation, *TRD* treatment-resistant depression, *WHO* World Health Organization

Table 2 Experts recruited and elicitation and aggregation method

Study	Experts and recruitment			Approach and method of elicitation			Aggregation method	
	Type of experts	Recruitment	Number of experts	VIM or FIM	Methods (summaries elicited)	Aggregation approach	Weights used in main exercise	Nature of weights
Cox et al. (2026) [43]	Clinicians with TMS expertise and clinical experience	Purposive sampling	7 (all 7 for operational parameters, 5 for TAU improvement parameters and 4 for long-term efficacy judgements)	VIM	Plausible lower and upper limits, fixed-interval probability allocation	Mathematical	NA	NA
Lee et al. (2025) [46]	Clinical oncologists (advanced RCC experts)	Purposive sampling	9	FIM	95% CIs	Mathematical	Yes	Equal weights
Lim et al. (2025) [47]	Clinicians (experts in managing pediatric and/or adult sepsis and/or experienced in antibiotic prescribing and regimen revision)	Purposive sampling	5	VIM	Mean, median, 25th percentile, 75th percentile, pooled cumulative distribution	Mathematical	Yes	Performance-based
Wang et al. (2025) [48]	Clinicians and clinical researchers including accredited exercise physiologists (oncology), physiotherapists, oncology nurses, clinical researchers in exercise oncology, and professorial researchers	Purposive sampling	12	VIM	2.5th percentile (L), 50th percentile (M), 97.5th percentile (U), mean and variance	Mathematical	Yes	Equal weights
Westwood et al. (2024) [23]	Clinical experts in emergency medicine, stroke, neurosurgery, and neuroradiology	Purposive sampling	5	FIM	Mode, upper, and lower bounds to fit a beta-PERT distribution	Mathematical	Yes	Equal weights

Table 2 (continued)

Study	Experts and recruitment			Approach and method of elicitation			Aggregation method	
	Type of experts	Recruitment	Number of experts	VIM or FIM	Methods (summaries elicited)	Aggregation approach	Weights used in main exercise	Nature of weights
Borsoi et al. (2023) [28]	Clinical experts in osteoporosis and REMS technology (gynecology, internal medicine, orthopedics and traumatology, endocrinology, rheumatology, and radiology)	Purposive sampling	6	VIM	Mode, minimum, maximum, and most likely: estimates checked for consistency	Mathematical	No	NA
Di Fusco et al. (2023) [27]	Clinicians, health economists, epidemiologists, policy strategists, and a patient representative	Purposive sampling	9	VIM	Mode/CI via Likert scales, prioritization of top value elements in Delphi process	Consensus	No	NA
Gorelova et al. (2023) [26]	Clinicians specializing in immunology and lab technicians familiar with DiGeorge syndrome	Purposive sampling	8 (4 clinicians, 4 lab technicians)	VIM	Probabilistic histograms and fixed 25% intervals for full distribution across possible values	Mathematical	No	NA
Gray et al. (2023) [25]	Clinical experts in glycemic management (pharmacists, endocrinologist, credentialled diabetes nurse educator, and quality-improvement staff)	Internal working group members	6	VIM	Mode, best estimate (most optimistic/pessimistic/realistic)	Consensus	No	NA
Squires et al. (2023) [24]	ED staff: two consultants, one nurse, and one junior doctor	Purposive sampling	4	FIM	Mean and variability (e.g., SD or range) for activity durations	Mathematical	No	NA
Willigers et al. (2023) [49]	Clinical experts (nephrologists) with international recognition in CKD research and clinical practice	Purposive sampling	6	FIM		Mathematical	Yes	Performance-based weights

Table 2 (continued)

Study	Experts and recruitment			Approach and method of elicitation			Aggregation method	
	Type of experts	Recruitment	Number of experts	VIM or FIM	Methods (summaries elicited)	Aggregation approach	Weights used in main exercise	Nature of weights
Ayers et al. (2022) [29]	Oncologists and hematologists	Purposive sampling	6	FIM	MLVs, LPLs, UPLs, CrIs, variance, CIs: expert interval width for truncated distribution variance	Consensus	No	NA
Paly et al. (2022) [44]	Clinical experts in advanced melanoma (oncologists)	Purposive sampling,	2	FIM	Lower bound, best estimate, upper bound of landmark survival (10 and 20 years)	N.R.	NA	NA
Klijn et al. (2021) [45]	Clinical expert with domain expertise in advanced renal cell carcinoma	Purposive sampling	1	VIM	Lower bound, upper bound, and most plausible value for long-term survival outcomes (10- and 20-year survival; mean survival)	N.A.	NA	NA
Konidaris et al. (2021) [31]	Clinicians with expertise in oncology, specifically for advanced CSCC treatment; health economists	Purposive sampling	NR	VIM	Mode, median, UPL, LPL, and CIs via SHELF guidance	Mathematical	No	NA
Petersohn et al. (2021) [30]	Vascular surgeons, cardiologists, internists, angiologists, PCPs, and vascular surgery nurses	Purposive sampling	NR	FIM	Upper/lower limits, mode for beta-PERT distribution	Mathematical	Yes	Equal weights
Sankatsing et al. (2020) [40]	Researchers with extensive breast cancer screening experience, involvement in pivotal trial	Purposive sampling	9 of 13 invited	NR	NR	Mathematical	No	NA
Stevenson et al. (2020) [33]	Clinical experts in neurodegenerative diseases such as CJD	Purposive sampling	4	FIM	Plausible lower and upper limits, median, IQR	Mathematical	No	NA

Table 2 (continued)

Study	Experts and recruitment			Approach and method of elicitation			Aggregation method	
	Type of experts	Recruitment	Number of experts	VIM or FIM	Methods (summaries elicited)	Aggregation approach	Weights used in main exercise	Nature of weights
Vokó et al. (2020) [32]	Experts in schizophrenia treatment and research	Purposive sampling	3	FIM	Median, IQR	Mathematical	Yes	Equal weights
Cope et al. (2019) [36]	Oncologists and hematologists with pediatric specialization and CAR-T therapy experience	Purposive sampling	7	FIM	MLV, LPL and UPL for survival rates at each time point	Mathematical	No	NA
Grand et al. (2019) [41]	Gynecologist specializing in endometriosis	Purposive sampling	1	FIM	NR	Mathematical	NR	NR
Hitimana et al. (2019) [35]	Gynecologists/obstetricians with ≥ 5 years of practice	Purposive sampling	8	VIM	Median and quartiles, optimistic/pessimistic scenarios, mode	Consensus	No	NA
Rossi et al. (2019) [34]	Urologists, radiologists, and oncologists with expertise in RCC diagnosis and management	Purposive sampling	4	FIM	Median, IQR, 95% CrIs	Mathematical	Yes	Equal weights
van Hoeven et al. (2019) [42]	Hematologists (adult and pediatric), neonatologists, immunologists	Purposive sampling	NR	FIM	Median; 95% CIs	Mathematical	Yes	Equal weights
Kip et al. (2018) [38]	Cardiologists with experience in NSTEMI diagnostic strategies and triple biomarker tests.	Purposive sampling	10	FIM	Event probabilities based on scenarios	Mathematical	No	NA
Rodriguez et al. (2018) [37]	Clinical experts in pediatric hemophilia	Purposive sampling	8	FIM	Mean, standard deviation, minimum, maximum	Consensus	No	NA
Grigore et al. (2017) [15]	PCPs based in the UK	Purposive sampling	18 (8 face-to-face, 7 remotely)	FIM	Upper/lower limits, mode for main question response	Mathematical	Yes	Equal weights

Table 2 (continued)

Study	Experts and recruitment			Approach and method of elicitation			Aggregation method	
	Type of experts	Recruitment	Number of experts	VIM or FIM	Methods (summaries elicited)	Aggregation approach	Weights used in main exercise	Nature of weights
Schnell-Inderst et al. (2017) [39]	Methodological experts (statisticians, epidemiologists, trialists) for internal biases; clinical experts (orthopedic surgeons) for external biases	Purposive sampling	9 methodologists; 11 orthopedists	FIM	Median, IQRs for beta distribution to reflect bias-adjustment	Mathematical	Yes	Equal weights

CAR-T chimeric antigen receptor T-cell (therapy), *CI* confidence interval, *CJD* Creutzfeldt Jakob disease, *CKD* chronic kidney disease, *CSCC* cutaneous squamous cell carcinoma, *CrI* (Bayesian) credible interval, *FIM* fixed interval method, *HTA* health technology assessment, *IQR* interquartile range, *L* low, *LPL* lower probability limit, *M* median, *MLV* most likely value, *NA* not applicable, *NR* not reported, *NSTEMI* non-ST elevation myocardial infarction, *PAD* peripheral artery disease, *PCP* primary care physician, *PERT* Program Evaluation and Review Technique, *RCC* renal cell carcinoma, *REMS* radiofrequency echographic multispectrometry, *SD* standard deviation, *SHELF* Sheffield Elicitation Framework, *TAU* treatment as usual, *TMS* transcranial magnetic stimulation, *U* upper, *UPL* upper probability limit, *VIM* variable interval method

reimbursement decisions, suggesting the particular value of expert elicitation in addressing uncertainty at earlier stages of health technology development. See Table 1 study-level details on the interventions, the timing of the assessment, and the parameter types elicited.

Experts consulted in these studies were in almost all cases clinicians, with specialty related to the targeted population and treatment evaluated (hematology/oncology, cardiology, neurology, radiology, and general practice). Recruitment used purposive sampling throughout, targeting clinicians with relevant specialization, though several studies incorporated multidisciplinary perspectives from health economists, epidemiologists, policy experts, and in one case, a patient representative [27]. Panel sizes ranged from a single to 18 experts, with a median of six experts per study. The encoding methods used were FIM (17 studies; 63%) and VIM (10 studies; 37%). The aggregation methods used in the studies were mainly mathematical (21 studies; 75%) or by consensus (6 studies; 21%), and one study did not require aggregation due to using a single expert. Notably, only 11 studies (39%) of studies explicitly reported using weighted aggregation, all except for two [47, 49] employed equal weights rather than performance-based or expertise-weighted schemes. See Table 2 for details on experts and elicitation and aggregation methods.

The conduct and analytical approaches demonstrated varying levels of methodological rigor, or at least variation in reporting and transparency. Elicitation formats consisted of face-to-face interviews, remote sessions via email or telephone, or web-based platforms, and these formats seemed to show a clear trend toward more remote and digital methods, with three quarters (21 studies; 75%) of the studies conducting elicitation entirely or partially through virtual participation. Training was provided to experts in two thirds of studies, typically covering the elicitation process, cognitive biases, and uncertainty quantification methods, while one third of the studies either did not provide training or did not report on this aspect. Pilot testing of elicitation materials was conducted in 20 studies (71%), suggesting recognition of the importance of refining elicitation protocols before the main exercise. Frameworks, protocols and tools included SHELF (7 studies; 25%), EXPLICIT (4 studies; 14%) , MRC (3 studies; 11%). Modified Delphi (2 studies; 7%) and Cook's method (1 study; 4%) [15, 22]. However, most studies (11 studies; 39%) did not report a framework. Many (20 studies; 71%) studies provided opportunities for experts to revise their initial estimates, either through iterative rounds or consensus discussions following individual elicitation. Mathematical aggregation, for the most part, used linear pooling; notably, with some using no pooling whatsoever and none using other methods such as e.g. log pooling. Distribution fitting methods included beta-PERT, parametric survival modeling, and least squares approaches.

Table 3 Conduct of the elicitation process and analysis

Study	Conduct					Analyses		
	Mode of administration	Opportunities for revision	Framework/protocol/tool	Training	Piloting	Pooling	Fitting	Pooled distribution used directly within DA model?
Cox et al. (2026) [43]	Individual, remote elicitation	No	MRC	Yes	Yes	Linear pooling	NR	Yes
Lee et al. (2025) [46]	Individual, remote elicitation	Yes	MRC	Yes	Yes	Linear pooling	Beta distributions fitted	Yes
Lim et al. (2025) [47]	Individual interviews, face-to-face versus remote not explicitly stated	No	MRC	Yes	No	Linear pooling	Dirichlet and beta distributions fitted	Yes
Wang et al. (2025) [48]	Individual remote elicitation	Yes	SHELF	Yes	Yes	Linear pooling	Beta distributions fitted	Yes
Westwood et al. (2024) [23]	Remote expert elicitation	Yes	EXPLICIT	Yes	No	Linear pooling	NR	Yes
Borsoi et al. (2023) [28]	Individual, remote administration	Yes	NR	No	Yes	Linear pooling	No formal statistical fitting applied	Yes
Di Fusco et al. (2023) [27]	Individual interviews for round 1; virtual panel discussions for rounds 2/3	Yes	Modified Delphi	Yes	Yes	No pooling	NR	No
Gorelova et al. (2023) [26]	Individual in-person interviews	Yes	NR	Yes	Yes	No pooling	No explicit fitting method was applied	No
Gray et al. (2023) [25]	In-person/remote interviews in round 1; remotely in session 2	Yes	NR	Yes	Yes	No pooling	NR	Yes
Squires et al. (2023) [24]	Remotely as a group	Yes	NR	Yes	Yes	Linear pooling	Gamma distribution fitted	Yes
Willigers et al. (2023) [49]	Individual remote elicitation	NR	Cooke's method	Yes	Yes	Linear pooling	Parametric survival models using several distributions	Yes
Ayers et al. (2022) [29]	Individual/remote interviews first; then remote consensus meeting	Yes	SHELF	Yes	Yes	Linear pooling	Yes, methods NR	Yes
Paly et al. (2022) [44]	Individual remote elicitation	Yes	SHELF	Yes	Yes	No pooling	Yes, methods NR	NR
Klijn et al. (2021) [45]	NR	NR	SHELF	NR	NR	NR	NR	NR
Konidaris et al. (2021) [31]	Individual and remote interviews	Yes	SHELF	Yes	Yes	No pooling	Parametric survival models, method NR	Yes
Petersohn et al. (2021) [30]	Remote elicitation	Yes	EXPLICIT	Yes	Yes	Linear pooling	Beta-PERT distribution fitted	Yes
Sankatsing et al. (2020) [40]	Online tool	NR	EXPLICIT	NR	NR	Linear pooling	Beta-PERT distributions	Yes

Table 3 (continued)

Study	Conduct					Analyses		
	Mode of administration	Opportunities for revision	Framework/protocol/tool	Training	Piloting	Pooling	Fitting	Pooled distribution used directly within DA model?
Stevenson et al. (2020) [33]	Face-to-face facilitated workshop using the SHELF framework.	Yes	SHELF	Yes	Yes	Consensus	Beta distribution fitted	Yes
Vokó et al. (2020) [32]	Individual face-to-face interviews	Yes	NR	Yes	Yes	Linear pooling	Dirichlet distribution fitted	Yes
Cope et al. (2019) [36]	Web-based application for remote elicitation	Yes	SHELF	Yes	Yes	Linear pooling	Parametric survival models combined expert estimates with RCT data	Yes
Grand et al. (2019) [41]	Web-based elicitation	NR	NR	NR	NR	No pooling	Yes, beta/Dirichlet distributions fitted	Yes
Hitimana et al. (2019) [35]	Individual face-to-face interviews; followed by remote follow-up	Yes	Modified Delphi	No	Yes	No pooling	No	Yes
Rossi et al. (2019) [34]	Face-to-face/phone interviews	Yes	NR and R	Yes	Yes	Linear pooling	Dirichlet and (modified) Connor–Mosimann distributions	Yes
van Hoeven et al. (2019) [42]	Online survey	Yes	NR	Yes	Yes	Linear pooling	Distributions via three-point estimation method	Yes
Kip et al. (2018) [38]	Remote online survey, followed consensus meeting	Yes	NR	No	No	No pooling	No	Yes
Rodriguez et al. (2018) [37]	Paper-based survey with closed-ended questions	No	NR	No	No	No pooling	No	Yes
Grigore et al. (2017) [15]	Separate facilitated face-to-face workshops for methodologists, clinicians	Yes	EXPLICIT	Yes	Yes	Linear pooling	NR	Yes
Schnell-Inderst et al. (2017) [39]	NR	NR	NR	NR	NR	Linear pooling	Least squares	Yes

CI confidence interval, *DA* decision-analytic, *EXPLICIT* Expert Elicitation Tool, *MRC* Medical Research Council (report, Bojke et al. 2021), *NR* not reported, *OS* overall survival, *PERT* Program Evaluation and Review Technique, *PSA* probabilistic sensitivity analysis, *RCC* renal cell carcinoma, *RCT* randomized controlled trial, *RR* relative risk, *SEE* structured expert elicitation, *SHELF* Sheffield Elicitation Framework, *STEER* Structured Expert Elicitation Resource

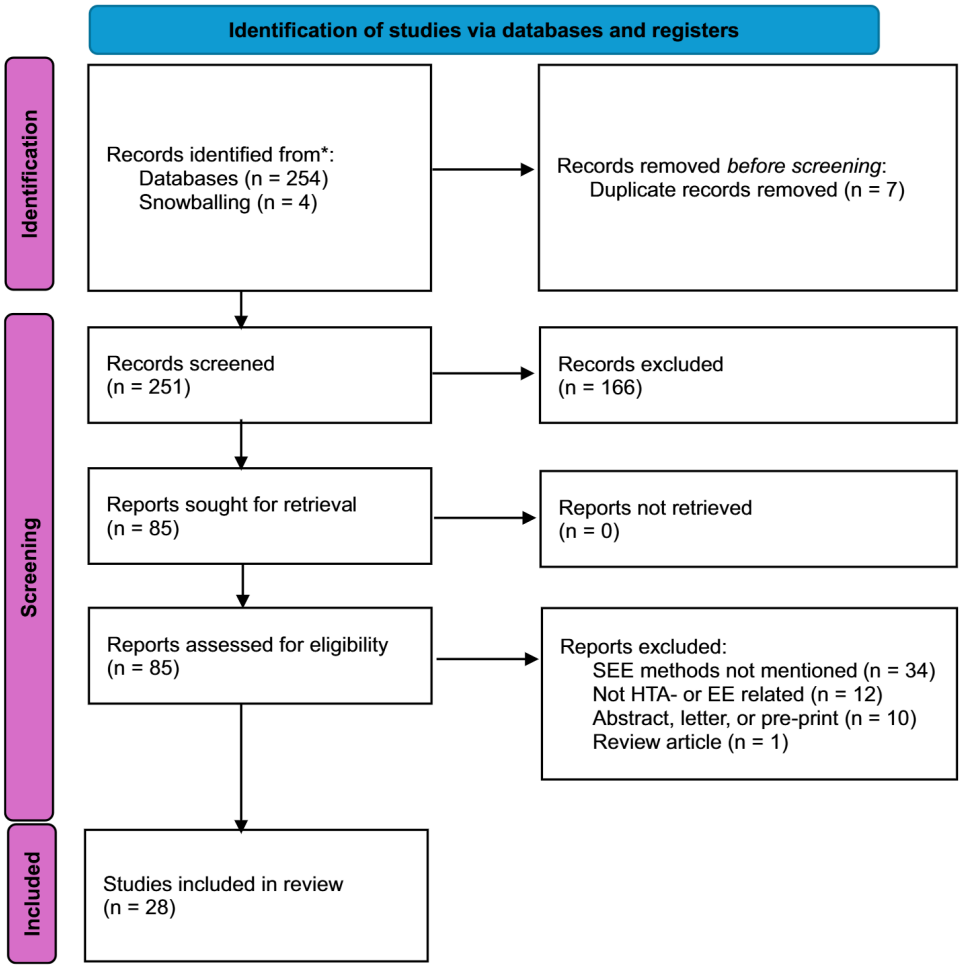

Fig. 1 PRISMA flow chart. The flow chart illustrates the study selection process according to PRISMA guidelines

For survival parameters, elicited estimates were sometimes combined with trial data using Bayesian approaches. Less commonly, cost and utility estimates were incorporated directly with minimal transformation. Overall, almost all studies (23 studies; 88% among the ones reporting) directly incorporated pooled expert distributions into decision models. See Table 3 for details on the conduct of the elicitation process and analysis.

4 Discussion

This systematic review identified nearly two-dozen studies demonstrating the growing but still limited application of structured expert elicitation in health economic evaluations, with a clear emphasis on addressing data scarcity in challenging contexts. We found that expert elicitation is predominantly employed for event and transition probabilities of treatment evaluations, particularly in rare diseases and specialized populations where traditional empirical evidence is insufficient, with nearly half of the studies supporting early-stage research and development decisions rather than formal value dossiers or HTA decisions. The methodological landscape shows standardization in some areas, with most studies consulting about seven clinical experts after training and piloting the process, though significant variation exists in the conduct and analysis. We also observed a trend toward more digital and remote elicitation methods.

The methodological approaches employed across the identified studies reveal separate important considerations regarding bias mitigation, expert variability, and external validity in SEE. Panel sizes demonstrated considerable variation, but the typical range of 6–8 experts aligns with recommendations for balancing diverse perspectives while maintaining manageable group dynamics [5]. Training provision in two thirds of the studies and opportunities for estimate revision in an even higher proportion suggests recognition of the importance of bias mitigation, though the lack of standardized training protocols across studies and variations in revision processes raises questions about the consistency and optimal application of these interventions.

The 3:1 predominance of mathematical aggregation methods over consensus approaches may either indicate a preference for preserving individual expert contributions rather

than requiring group agreement, though this choice carries implications for how between-expert variability is handled and whether minority opinions are adequately captured, or may simply be more convenient. However, relatively few studies explicitly reported using weighted aggregation approaches, and notably, of those that did, all employed equal weighting schemes rather than performance-based or expertise-weighted methods, which could take expert credibility or track records into account. This could be due to multiple reasons, but it takes time to allocate proper weights, and this is not an easy task, so both feasibility and methodological concern are likely reasons. While the MRC protocol [8] notes that the literature on pooling methods is not sufficiently mature to recommend alternatives to linear pooling, it also suggests that there may nonetheless be valid reasons to consider different aggregation methods depending on the decision context [8]. The universal reliance on linear pooling in identified studies is thus consistent with current guidance, though the comparative performance of alternative approaches remains an area warranting further investigation.

The choice between variable and fixed interval methods showed—contrary to the findings by Soares et al.—a potential preference for FIM over VIM, which may reflect practical considerations around elicitation complexity and expert burden [10, 18]. The choice between VIM and FIM may depend on the specific elicitation context, and there is no definite evidence to prefer either over the other. Guidelines such as the MRC protocol and NICE Technical Support Document (TSD) 26 notes several dimensions for analysts to be aware of when selecting either method, and while the ISPOR recommendations suggest that the MRC protocol has a slight preference for FIM, the NICE TSD 26 has a cautions recommendation for considering VIM first [5, 8, 10]. However, these recommendations seem more related to guiding analysts to avoid potential issues with FIM, rather than VIM being methodologically superior in general. This context-specificity suggests that guidance on encoding methods should consider the intended analytical use of elicited quantities rather than advocating for one approach universally.

The rate of publication—28 studies over a 9-year period—is somewhat lower compared with the prior review by Soares et al., which reported on the same number of studies but over a period of only 4 years [5, 8, 9, 50]. The lack of an observable increase in publication rates may partly reflect the substantial lag between the publication of methodological guidance and its manifestation in peer-reviewed applications. Given that NICE's explicit mention of SEE in the 2022 Methods Manual and subsequent detailed guidance in TSD 26 are relatively recent, the expected increase in methodologically rigorous SEE applications may not yet be captured in peer-reviewed literature.[9, 10] Based on typical timelines for health economic research, guidance uptake might reasonably be expected to become apparent in publications from 2025 to 2027 onward. Importantly, the rate of peer-reviewed publication may not accurately reflect the actual uptake of SEE methods in practice. A substantial proportion of SEE applications occur within internal early-stage modelling for product development decisions or within HTA submissions that may never be published in the peer-reviewed literature. Consequently, our findings likely underestimate the true extent of SEE use in health economic decision-making.

Generally, there is a substantial proportion of "not reported" entries across Tables 1–3, which highlights persistent reporting deficiencies in SEE applications. Key methodological details including weighting approaches, distribution fitting methods, and piloting procedures were frequently omitted, limiting the ability to assess methodological rigor or replicate approaches. This underreporting underscores the need for standardized reporting guidelines such as Structured Expert Elicitation Resource (STEER) to improve transparency and enable more robust evidence synthesis [11].

The application patterns observed in this review highlight the strategic deployment of SEE in addressing specific evidence gaps within health economic evaluation. The concentration on treatment evaluations, particularly in rare diseases and specialized populations, demonstrates the practical value of expert knowledge when traditional empirical evidence is insufficient or unavailable. As rare diseases have received more attention from research and development in the pharmaceutical and biotech industries, we find it somewhat puzzling why there are not more SEE application in the more recent past [51]. A reason for that is perhaps that such quantitative and explicit use of SEE is still in its infancy. The uptake may be slowed by HTA bodies' guidelines only recently starting to include such methodologies. It may also be due to SEE somewhat being perceived as challenging the evidence-based medicine (EBM) and the randomized controlled trial (RCT) traditions [1]. Additionally, it is possible that SEE methods are being used more frequently but are either not being reported in peer-reviewed publications or described without enough methodological detail to be identified through systematic database searches. This underreporting may reflect that SEE could be embedded within broader HTA submissions or internal decision-making processes where detailed methodological descriptions are not prioritized. For example, Forsyth et al. conducted a systematic review of SEE for long-term survival outcomes in HTA reports of oncology treatments for NICE [52]. They found that 5/10 studies could be identified via database searches, whereas 4/10 were found in HTA reports. The finding that nearly half of studies supported early-stage research and development decisions rather than formal reimbursement assessments may be a nod toward the particular utility of expert elicitation during technology development phases, where rapid

decision-making is often required before comprehensive clinical evidence becomes available. Using expert opinion and SEE for critical parts of decision-analytic models that lead to early access via managed entry agreements would require that additional, more robust evidence will become available at a later stage to re-evaluate reimbursement agreements. The focus on event and transition probabilities as the most frequently elicited parameters reflects the critical importance of expert knowledge in quantifying clinical pathways and disease progression and areas where observational data may be particularly limited or where extrapolation beyond trial populations is necessary.

The heterogeneity in methodological approaches observed across studies points to several priority areas for future research and standardization efforts. The development of more rigorous guidelines for expert selection, panel size determination, and weighting schemes represents a critical need, particularly given the limited use of performance-based weighting and the wide variation in panel compositions. It will also be important with improvements in the standardization of reporting to ensure future SLRs can synthesize even more information currently possible. Future research, and subsequently guidelines such as ISPOR and other methodological frameworks, should address the choice between mathematical and consensus-based aggregation methods, providing clearer guidance on when each approach is most appropriate based on the decision context and parameter characteristics [5, 8–10, 50]. The emerging trend toward remote and digital elicitation methods—possibly ignited by the coronavirus disease (COVID-19) pandemic, which fell into our catchment period, but which seems to continue—necessitates research into the comparative validity and reliability of different elicitation formats, particularly as these approaches become more prevalent in postpandemic research environments. A valuable future research opportunity lies in validating the accuracy of expert predictions through comparison with subsequently available empirical data. This will require effort, perhaps even stated requirements, to evaluate and compare the precision of SEE with future longer follow-up data from clinical trials, real-world data (RWD), and other postmarket evidence. Such validation requirements would be most applicable to SEE applications supporting regulatory decisions or managed access arrangements, where subsequent evidence generation is typically expected, rather than to internal early-stage modelling where such follow-up may not occur.

This may be of significant value to increase acceptance and use with HTA bodies and decision-makers. Our recommendations for future research align substantially with those identified in the MRC protocol and ISPOR guidance, reinforcing the continued relevance of these priority areas and suggesting limited progress in addressing them through empirical applications to date [5, 8].

Standardization efforts should also focus on developing more consistent training protocols and bias mitigation strategies, i.e., in both individual and consensus elicitation. The integration of structured expert elicitation with existing health technology assessment processes will benefit from further methodological development, including guidelines for transparency in reporting, validation approaches for expert-derived parameters, and frameworks for sensitivity analysis around expert-informed inputs. Additionally, future research could continue to explore the relevance of expertise-weighted aggregation methods and should investigate the comparative performance of different pooling approaches, particularly in contexts where expert disagreement is substantial or where parameter uncertainty has significant decision implications.

Our work is subject to several limitations. First, our review was restricted to peer-reviewed journal articles, potentially missing relevant applications of SEE in grey literature or HTA reports not published as peer-reviewed manuscripts in journals. However, our objective was to provide a review update of the peer-reviewed literature, and future studies can choose a broader SCOPR. Second, our search strategy did not include specific protocol names (e.g., "SHELF" or "MRC protocol") as search terms, which may have resulted in missed studies that referenced these frameworks without using broader expert elicitation terminology. However, future reviews may benefit from including protocol-specific search terms. Third, the heterogeneity in reporting styles and level of detail across studies made it challenging to consistently extract and compare information about SEE processes. However, this variability also highlights the need for standardized reporting guidelines in SEE applications, which could be a valuable outcome of our review. Fourth, our search strategy, while comprehensive, may have missed some relevant studies due to the diverse terminology used in the field of expert elicitation. However, our use of multiple databases and snowball sampling techniques likely mitigated this risk to a significant extent. Finally, our review did not include a formal quality assessment of the included studies, which could have provided additional insights into the methodological rigor of SEE applications. However, the detailed extraction and analysis of methodological approaches used in each study offer a comprehensive overview of current practices and their evolution over time.

5 Conclusions

The landscape of SEE in economic evaluations and HTA over the past 8 years is evolving although publication activity has not accelerated. Methods still vary considerably across SEE applications, suggesting that the optimal approaches have yet to emerge across several key dimensions

of expert elicitation. Nevertheless, there seems to be some recognition of SEE's potential in informing healthcare decision-making, particularly in rare diseases and subpopulations of larger disease entities where empirical data are scarce or even unavailable. As SEE assumes an increasingly prominent role in HTA processes, future research should prioritize standardization of best practices, validation of expert predictions against subsequently available empirical data, and development of more bias mitigation strategies. Such advances will be essential for enhancing the credibility and reliability of expert-informed analyses in health policy and reimbursement decisions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40273-026-01628-x.

Acknowledgements The co-authors would like to acknowledge Yancy S. Wu for his input in the earlier work related to this review.

Funding Open access funding provided by University of Oslo (incl Oslo University Hospital). B.P.G. and E.A. receive funding for this work from the ASCERTAIN project (Grant Agreement 101094938), funded by the European Union, and the Research Council of Norway as part of NORCHER (Research Grant no. 296114). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA).

Declarations

Conflict of interest Fredrik Holmboe is an employee of AbbVie and may hold AbbVie stock. None of the other co-authors declares a financial or nonfinancial interest that is directly or indirectly related to this work.

Data Availability Statement Files with references and extraction tables can be made available to interested parties upon reasonable request.

Author Contributions Conceptualization: E.A. Methodology: B.P.G.. Investigation: B.P.G., I.S., and F.H. Data curation: B.P.G. and I.S. Formal analysis: all co-authors. Writing—original draft: B.P.G. Writing—review and editing: all co-authors. Supervision: E.A. Funding acquisition: E.A.

References

1. Morgan MG. Use (and abuse) of expert elicitation in support of decision-making for public policy. Proc Natl Acad Sci. 2014;111(20):7176–84.
2. Dias LC, Morton A, Quigley J. Elicitation: the science and art of structuring judgement, 1st ed. Cham: Springer Nature; 2018. https://doi.org/10.1007/978-3-319-65052-4
3. O'Hagan A. Uncertain judgements: eliciting experts' probabilities. London: Wiley; 2006.
4. O'Hagan A. Expert knowledge elicitation: subjective but scientific. Am Stat. 2019;73(sup1):69–81.
5. Soares M, Colson A, Bojke L, Ghabri S, Garay OU, Felli JK, et al. Recommendations on the use of structured expert elicitation protocols for healthcare decision making: a good practices report of an ISPOR task force. Value in Health. 2024;27(11):1469–78.
6. Morris DE, Oakley JE, Crowe JA. A web-based tool for eliciting probability distributions from experts. Environ Modell Softw. 2014;52:1–4.
7. Bojke L, Grigore B, Jankovic D, Peters J, Soares M, Stein K. Informing reimbursement decisions using cost-effectiveness modelling: a guide to the process of generating elicited priors to capture model uncertainties. Pharmacoeconomics. 2017;35(9):867–77.
8. Bojke L, Soares M, Claxton K, Colson A, Fox A, Jackson C, et al. Developing a reference protocol for structured expert elicitation in healthcare decision-making: a mixed-methods study. Health Technol Assess. 2021 Jun;25(37):1–124.
9. National Institute for Health and Care Excellence. NICE health technology evaluations: the manual. 2022. Available from: https://www.nice.org.uk/process/pmg36. Accessed 29 May 2026.
10. Oakley JE, Ren2 S, Forsyth JE, Gosling JP, Wilson K, Latimer N, et al. TSD 26: Expert elicitation for long-term survival outcomes. 2025. Available from: https://sheffield.ac.uk/media/94031/download?attachment. Accessed 29 May 2026.
11. Jankovic D, Horscroft J, Lee D, Bojke L, Soares M. STEER: open access resources for conducting structured expert elicitation for health care decision making. Med Decis Making. 2025;45(6):627–39.
12. Holzhauer B, Hampson LV, Gosling JP, Bornkamp B, Kahn J, Lange MR, et al. Eliciting judgements about dependent quantities of interest: the Sheffield Elicitation Framework extension and copula methods illustrated using an asthma case study. Pharm Stat. 2022;21(5):1005–21.
13. Hemming V, Burgman MA, Hanea AM, McBride MF, Wintle BC. A practical guide to structured expert elicitation using the IDEA protocol. Methods Ecol Evol. 2018;9(1):169–80.
14. Cooke RM. Experts in uncertainty: opinion and subjective probability in science. New York: Oxford University Press; 1991.
15. Grigore B, Peters J, Hyde C, Stein K. EXPLICIT: a feasibility study of remote expert elicitation in health technology assessment. BMC Med Inform Decis Mak. 2017;17(1):131.
16. Riley D, Fox A, Hirst A, Marciniak A, Heron L. Expert elicitation techniques: informing application in health technology assessment decision making. Value Outcomes Spotlight. 2025;11(4):40–3.
17. Grigore B, Peters J, Hyde C, Stein K. Methods to elicit probability distributions from experts: a systematic review of reported practice in health technology assessment. Pharmacoeconomics. 2013;31(11):991–1003.
18. Soares MO, Sharples L, Morton A, Claxton K, Bojke L. Experiences of structured elicitation for model-based cost-effectiveness analyses. Value in Health. 2018;21(6):715–23.
19. Cadham CJ, Knoll M, Sánchez-Romero LM, Cummings KM, Douglas CE, Liber A, et al. The use of expert elicitation among computational modeling studies in health research: a systematic review. Med Decis Making. 2022;42(5):684–703.

20. Forsyth JE, Uttley L, Wong R, Ren S. Structured expert elicitation for long-term survival outcomes in health technology assessment: a systematic review. BMC Med Inform Decis Mak. 2025;25(1):387.
21. Jandhyala R. Delphi, non-RAND modified Delphi, RAND/UCLA appropriateness method and a novel group awareness and consensus methodology for consensus measurement: a systematic literature review. Curr Med Res Opin. 2020;36(11):1873–87.
22. Gosling JP. SHELF: the Sheffield Elicitation Framework. In: Dias LC, Morton A, Quigley J, editors. Elicitation: the science and art of structuring judgement. Cham: Springer International Publishing; 2018. p. 61–93.
23. Westwood M, Ramaekers B, Grimm S, Armstrong N, Wijnen B, Ahmadu C, et al. Software with artificial intelligence-derived algorithms for analysing CT brain scans in people with a suspected acute stroke: a systematic review and cost-effectiveness analysis. Health Technol Assess (Winchester, England). 2024;28(11):1–204.
24. Squires H, Mason S, O'Keeffe C, Croft S, Millington G. What impact would reducing low-acuity attendance have on emergency department length of stay? A discrete event simulation modelling study. Emerg Med J. 2023;41(1):27–33.
25. Gray J, Thynne TR, Eaton V, Reade B, Larcombe R, Baldacchino L, et al. A framework for local-level economic evaluation to inform implementation decisions: health service interventions to prevent hospital-acquired hypoglycemia. Int J Technol Assess Health Care. 2023;39(1):e74.
26. Gorelova M, Rysankova K, Donin G, Kneppo P, Rogalewicz V. Comparison of elicitation approaches in early stage HTA applied on artificial thymus for patients with DiGeorge syndrome. Healthcare. 2023;11(22):20.
27. Di Fusco M, Mendes D, Steuten L, Bloom DE, Drummond M, Hauck K, et al. The societal value of vaccines: expert-based conceptual framework and methods using COVID-19 vaccines as a case study. Vaccines. 2023;11(2):234. https://doi.org/10.3390/vaccines11020234.
28. Borsoi L, Armeni P, Brandi ML. Cost-minimization analysis to support the HTA of radiofrequency echographic multi spectrometry (REMS) in the diagnosis of osteoporosis. Glob Reg Health Technol Assess. 2023;10(1):1–11.
29. Ayers D, Cope S, Towle K, Mojebi A, Marshall T, Dhanda D. Structured expert elicitation to inform long-term survival extrapolations using alternative parametric distributions: a case study of CAR T therapy for relapsed/ refractory multiple myeloma. BMC Med Res Methodol. 2022;22(1):272.
30. Petersohn S, Grimm SE, Ramaekers BLT, ten Cate-Hoek AJ, Joore MA. Exploring the feasibility of comprehensive uncertainty assessment in health economic modeling: a case study. Value Health. 2021;24(7):983–94.
31. Konidaris G, Paul E, Kuznik A, Keeping S, Chen CI, Sasane M, et al. Assessing the value of cemiplimab for adults with advanced cutaneous squamous cell carcinoma: a cost-effectiveness analysis. Value in Health. 2021;24(3):377–87.
32. Voko Z, Bitter I, Mersich B, Rethelyi J, Molnar A, Pitter JG, et al. Using informative prior based on expert opinion in Bayesian estimation of the transition probability matrix in Markov modelling—an example from the cost-effectiveness analysis of the treatment of patients with predominantly negative symptoms of schizophrenia with cariprazine. Cost Effect Resource Alloc. 2020;18(1):28. https://doi.org/10.1186/s12962-020-00224-w.
33. Stevenson M, Uttley L, Oakley JE, Carroll C, Chick SE, Wong R. Interventions to reduce the risk of surgically transmitted Creutzfeldt-Jakob disease: a cost-effective modelling review. Health Technol Assess. 2020;24(11):1–151.
34. Rossi SH, Blick C, Nathan P, Nicol D, Stewart GD, Wilson ECF. Expert elicitation to inform a cost-effectiveness analysis of screening for renal cancer. Value Health. 2019;22(9):981–7.
35. Hitimana R, Lindholm L, Mogren I, Krantz G, Nzayirambaho M, Sengoma JPS, et al. Incremental cost and health gains of the 2016 WHO antenatal care recommendations for Rwanda: results from expert elicitation. Health Res Policy Syst. 2019;17(1):36. https://doi.org/10.1186/s12961-019-0439-9.
36. Cope S, Ayers D, Zhang J, Batt K, Jansen JP. Integrating expert opinion with clinical trial data to extrapolate long-term survival: a case study of CAR-T therapy for children and young adults with relapsed or refractory acute lymphoblastic leukemia. BMC Med Res Methodol. 2019;19(1):182.
37. Rodriguez-Zepeda MDC, Gonzalez L, Bravo A, Pompa T, Silva S, Paredes R, et al. Cost-effectiveness of rFVIIa versus pd-aPCC in the management of mild to moderate bleeds in pediatric patients with hemophilia A with inhibitors in Mexico. Value in Health Regional Issues. 2018;17:164–73.
38. Kip MMA, Steuten LMG, Koffijberg H, Ij MJ, Kusters R. Using expert elicitation to estimate the potential impact of improved diagnostic performance of laboratory tests: a case study on rapid discharge of suspected non-ST elevation myocardial infarction patients. J Eval Clin Pract. 2018;24(1):31–41.
39. Schnell-Inderst P, Iglesias CP, Arvandi MARJAN, Ciani ORIANA, Matteucci Gothe R, Peters J, et al. A bias-adjusted evidence synthesis of RCT and observational data: the case of total hip replacement. Health Econ. 2017;26(Supplement 1):46–69.
40. Sankatsing VDV, Juraniec K, Grimm SE, Joore MA, Pijnappel RM, de Koning HJ, et al. Cost-effectiveness of digital breast tomosynthesis in population-based breast cancer screening: a probabilistic sensitivity analysis. Radiology. 2020;297(1):40–8.
41. Grand TS, Basarir H, Jackson LJ. The cost-effectiveness of oral contraceptives compared to "no hormonal treatment" for endometriosis-related pain: an economic evaluation. PLoS ONE. 2019;14(1):e0210089.
42. van Hoeven LR, Janssen MP, Lieshout-Krikke RW, Molenaar-de Backer MW. An assessment of the risk, cost-effectiveness, and perceived benefits of anti-parvovirus B19 tested blood products. Transfusion. 2019;59(7):2352–60.
43. Cox E, Ma J, Roadevin C, Morriss R, James M. Is it time for mental health services to invest in neurostimulation? An economic evaluation of transcranial magnetic stimulation therapies for the treatment of moderate to severe treatment-resistant depression in the UK. BMJ Ment Health. 2026;29(1):e302237.
44. Federico Paly V, Kurt M, Zhang L, Butler MO, Michielin O, Amadi A, et al. Heterogeneity in survival with immune checkpoint inhibitors and its implications for survival extrapolations: a case study in advanced melanoma. MDM Policy Pract. 2022;7(1):23814683221089659.
45. Klijn SL, Fenwick E, Kroep S, Johannesen K, Malcolm B, Kurt M, et al. What did time tell us? A comparison and retrospective validation of different survival extrapolation methods for immuno-oncologic therapy in advanced or metastatic renal cell carcinoma. Pharmacoeconomics. 2021;39(3):345–56.
46. Lee D, Ahmad Z, Larkin JMG, Bahl A, Melendez-Torres GJ. Structured expert elicitation to inform long-term survival extrapolations in advanced renal cell carcinoma. Appl Health Econ Health Policy. 2025;23(6):1073–83.
47. Lim C, Maung Swe MM, Devine A, Oakley T, Champlin K, Sone Oo P, et al. Cost-effectiveness of maintaining an active hospital microbiology laboratory service in Timor-Leste. Lancet Reg Health. 2025. https://doi.org/10.1016/j.lansea.2025.100582.
48. Wang Y, McCarthy AL, Tuffaha H. Co-designing a structured expert elicitation with clinicians to enhance health care decision making in exercise oncology. Med Decis Making. 2025;45(5):602–13.
49. Willigers BJA, Ouwens M, Briggs A, Heerspink HJL, Pollock C, Pecoits-Filho R, et al. The role of expert opinion in projecting long-term survival outcomes beyond the horizon of a clinical trial. Adv Ther. 2023;40(6):2741–51.
50. Peel A, Jenks M, Choudhury M, Lovett R, Rejon-Parrilla JC, Sims A, et al. Use of expert judgement across nice guidance-making programmes: a review of current processes and suitability of existing

tools to support the use of expert elicitation. Appl Health Econ Health Policy. 2018;16(6):819–36.
51. Citeline. Rare disease R&D: continued growth amid challenges. 2025 12/21/2025]. Available from: https://www.citeline.com/en/resources/rare-disease-r-and-d. Accessed 29 May 2026.
52. Alarid-Escudero F, Knowlton G, Easterly CW, Garibay D, Enns E. dampack: decision-analytic modeling package. 2024. Available from: https://cran.r-project.org/web/packages/dampack/index.html. Accessed 29 May 2026.