International Journal of Technology Assessment in Health Care

www.cambridge.org/thc

Method

Cite this article: Dóczy V, Wernerné Sódar B, Hölgyesi Á, Merész G, Gaál P (2022). Development, testing, and implementation of a new procedure to assess the clinical added benefit of pharmaceuticals. *International Journal of Technology Assessment in Health Care*, **38**(1), e58, 1–8
https://doi.org/10.1017/S0266462322000411

Received: 01 March 2022
Revised: 13 May 2022
Accepted: 01 June 2022

Key words:
Technology assessment; Health (MeSH Unique ID: D013673); Medical oncology (MeSH Unique ID: D008495); Comparative effectiveness research (MeSH Unique ID: D057186); Outcome assessment; health care (MeSH Unique ID: D017063); Decision making; organizational (MeSH Unique ID: D003659)

Author for correspondence:
*Veronika Dóczy,
E-mail: doczy.veronika@ogyei.gov.hu

The authors thank Professor Nathan Cherny and Mrs. Nicola Jane Latino and all the experts at ESMO-MCBS helpdesk for their quick responses and irreplaceable support, all stakeholders who provided feedback and thus improved the framework, Ms. Szilvia Szabó and all other clinical assessors of NIPN for their tireless help with the retrospective testing, the Department of Pharmacovigilance at NIPN for their useful contribution, and Ms. Antonia Bosworth Smith and Mr. Kenneth Paul White for their contribution to language editing.

© The Author(s), 2022. Published by Cambridge University Press.

CAMBRIDGE UNIVERSITY PRESS

Development, testing, and implementation of a new procedure to assess the clinical added benefit of pharmaceuticals

Veronika Dóczy[1*], Barbara Wernerné Sódar[1], Áron Hölgyesi[1,2], Gergő Merész[1,3] and Péter Gaál[4,5]

[1]Department of Health Technology Assessment, National Institute of Pharmacy and Nutrition, Budapest, Hungary; [2]Doctoral School of Molecular Medicine, Semmelweis University, Budapest, Hungary; [3]Doctoral School of Mental Health Sciences, Semmelweis University, Budapest, Hungary; [4]Faculty of Health and Public Administration, Health Services Management Training Centre, Semmelweis University, Budapest, Hungary and [5]Faculty of Technical and Human Sciences, Department of Applied Social Sciences, Sapientia Hungarian University of Transylvania, Targu Mures, Romania

Abstract

Objectives: The reimbursement process for innovative health technologies in Hungary lacks any formalized assessment of clinical added benefit (CAB). The aim of this research is to present the development, retrospective testing, and implementation of a local assessment framework for determining the CAB of cancer treatments at the Department of Health Technology Assessment of the National Institute of Pharmacy and Nutrition in Hungary.

Methods: The assessment framework was drafted after screening existing methods and a retrospective comparison of local reimbursement dossiers to that of German and French methods. The Magnitude of Clinical Benefit Scale of the European Society for Medical Oncology was chosen to rate the extent of CAB in oncology, as part of a conclusion complemented by the assessment of endpoint relevance and the quality of evidence. Several rounds of retrospective assessments have been conducted involving all clinical assessors, iterated with semistructured discussions to consolidate divergence between assessors. External stakeholders were consulted to provide feedback on the framework.

Results: Retrospective assessments resulted in average more than 75 percent concordance between assessors on each element of the conclusion. Input from ten stakeholders was also incorporated; stakeholders were generally supportive, and they mostly commented on the concept, the elements of the framework, and its implementation.

Conclusions: The procedure is suitable for routine use in the decision-making process to describe the CAB of antineoplastic technologies in Hungary. Further extension of the framework is required to cover more disease areas for structured and comparable conclusions on CAB of innovative health technologies.

Health technology assessment (HTA) is an evidence-based scientific method used to assess the added value, and ultimately, the reimbursement of innovative health technologies. In order to help local decision making, some national HTA bodies provide a conclusion on the clinical added benefit (CAB) of health technologies in their remit (1). However, the exact procedure of formulating a conclusion on CAB is unique to each setting, as policy goals, methodological guidelines, capacities, and other technical circumstances may differ (2). European HTA bodies, such as the ones operating in Germany (*Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen*—IQWiG) (3) or France (*Haute Autorité de Santé*—HAS) (4) have developed their own classification systems to assess CAB; other countries like Hungary have not done so. The reason for this delay might be related to the legal framework and resource constraints of HTA in the Central and Eastern European region, which has been discussed by other researchers (5;6).

Although frameworks already exist for making recommendations in clinical practice (e.g., GRADE) or in reimbursement decision making (the ones used by HAS or IQWiG), their modification is not advised (7). Adopting other agencies' classification is also not trivial as according to Boucaud-Maitre et al. (8), there were not more than 50 percent concordance in ratings between HAS and IQWiG, discrepancies potentially caused by differences in choosing the locally appropriate comparators and target populations. This raises the need for the development of a new, tailored framework, which could facilitate evidence-based reimbursement, in adherence with tasks, timelines, the local legal environment, and resource constraints in the daily routine of Hungarian HTA. Moreover, the legislation of the European Commission on joint clinical assessments for health technologies envisages an environment where sufficient information is available to conclude on the added benefit (9).

We propose a procedure embracing the Magnitude of Clinical Benefit Scale developed by the European Society for Medical Oncology (ESMO-MCBS) as a core component to assess the extent

of CAB for antineoplastic medicines. The ESMO-MCBS scores drugs based on their impact on survival (baseline score) but enables upgrading/downgrading based on toxicity and quality of life outcomes (adjusted score) into a three-point (A–C) or a five-point (1–5) scale depending on curative and noncurative settings, respectively (10).

The aim of this research is to present the development, pilot testing, stakeholder consultation, and implementation of the proposed framework for assessing the CAB of antineoplastic drugs at the Department of HTA of the National Institute of Pharmacy and Nutrition (NIPN).

Methods

The design process of the framework can be divided into the steps of drafting, testing, feedback assessment from stakeholders, and implementation. Internal and external phases can be distinguished by whether the contributors of the steps were employees of NIPN or other stakeholders are involved (Figure 1).

Review and Retrospective Analyses of Local Reimbursement Dossiers

First, two medical assessors screened the document repositories of European HTA agencies for current practices on CAB assessment. We concluded on using the Transparency Committee Doctrine of HAS (4) and the General methods v6.1 of IQWiG (3) as a basis for the development of our procedure.

In order to investigate whether the distribution of CAB categories of dossiers submitted to NIPN was similar to those evaluated by HAS and IQWiG, retrospective analysis was performed on all reimbursement dossiers (irrespective of their indication) submitted to the Department of HTA in 2019 and 2020. The corresponding clinical added value (CAV) was retrieved from the assessments published on HAS's website. Categories CAV I–II were merged as "Major added value." This allowed us to perform a comparison with the distributions of CAV categories presented in a recently published study (8) comparing the scoring systems of IQWiG and HAS. Next, the reimbursement submissions were categorized based on the authorized therapeutic indications.

Internal Development

A draft framework was proposed and then further elaborated in two rounds of internal discussions, with emphasis on the feasibility of the proposed procedure.

Retrospective Testing

The retrospective testing and the implementation of the framework were done in four consecutive rounds. First, the matured draft version of the procedure was piloted by two previously uninvolved assessors, and their feedback was also discussed and incorporated into the documentation. They evaluated two current reimbursement dossiers using ESMO-MCBS and compared them to the published score available at the ESMO website (round #1). Additional rounds (#2–#4) of retrospective assessments involved all available clinical assessors ($n = 7$ or 9) to identify any divergence and to build consensus on handling such cases.

Feedback from Stakeholders

A call for open consultation and a working paper on CAB was posted on the website of NIPN on 15 June 2021, with a deadline for feedback

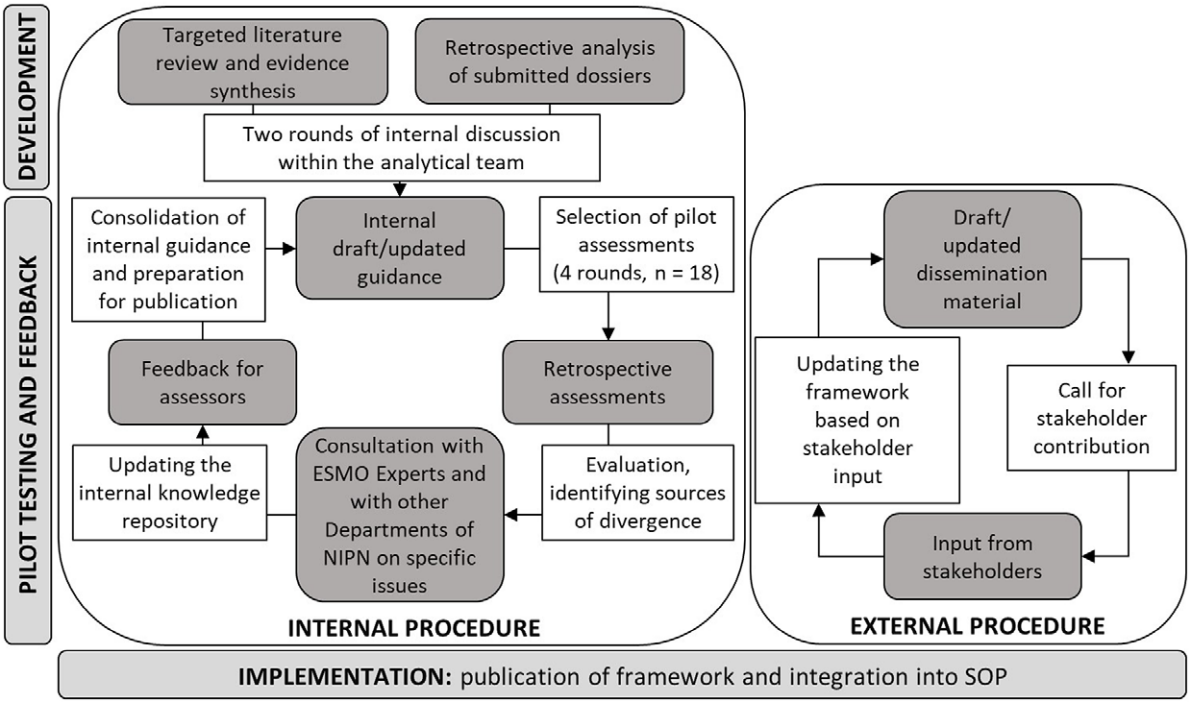

Figure 1. The schematic representation of the framework development.
The designing of the framework can be divided into the steps of development, retrospective testing and feedback, and implementation; we can also distinguish internal and external phases, depending on whether the contributions were made solely by NIPN employees or if other stakeholders were involved.
Notes. Some retrospective test and the stakeholder consultation were performed simultaneously, for time-saving purposes.
CAB, clinical added benefit; NIPN, National Institute of Pharmacy and Nutrition; SOP, standard operating procedure.

https://doi.org/10.1017/S0266462322000411 Published online by Cambridge University Press

Table 1. Comparison of the Distribution of CAB/CAV Categories Across Dossiers Submitted to HAS, IQWiG, and NIPN

NIPN retrospectively assigned CAV levels retrieved from HAS (2019–2020) $n = 213$ (102 in 2019 and 111 in 2020)			CAV categories reported by Boucaud-Maitre et al, 2020			
			HAS (2011–2017)		IQWiG (2011–2017)	
Category	*n* (%)	% of evaluated by HAS	Category	*n* (%)	Category	*n* (%)
CAV I/II (major/important)	4 (1.9)	2.4	CAV I/II (major/important)	2 (1.0)	Major	20 (10.5)
CAV III (moderate)	25 (11.7)	15.2	CAV III (moderate)	31 (16.2)	Considerable	32 (16.7)
CAV IV (minor)	43 (20.2)	26.2	CAV IV (minor)	57 (29.8)	Minor	16 (8.4)
CAV V (no improvement)	75 (35.2)	45.7	CAV V(no improvement) and CAV insufficient*	101 (52.9)	No proof of benefit	106 (55.5)
CAV insufficient	17 (8.0)	10.4			Not quantifiable	15 (7.9)
Not evaluated by HAS	49 (23.0)					
Total	213 (100.0)	100.0	Total	191 (100.0)	Total	191 (100.0)

Notes. 2011–2017 HAS and IQWiG classifications were adapted from Boucaud-Maitre et al, 2020, in the case of NIPN 2019–2020 we present the retrospectively assigned HAS classification categories retrieved from HAS's website. * these two categories are merged in Boucaud-Maitre et al, 2020.
CAB, clinical added benefit; CAV, clinical added value; HAS, Haute Autorité de Santé; IQWiG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; *n*, number of assessed dossiers; NIPN, National Institute of Pharmacy and Nutrition.

on 17 September 2021. Dissemination materials briefly described the procedure of formulating the conclusion on the CAB in general, and its potential impact on the reimbursement process. There was no restriction on who could reply to the call, but twenty-four entities (patient organizations, medical societies, academic centers, public bodies, industry associations, and consultancy firms) were invited to comment on the working paper. A self-administered questionnaire was provided for stakeholders in which respondents could express their response on a four-level Likert scale to a predefined set of questions on four different domains (Supplementary Table 2). Their feedback was also gathered as comments without restrictions via a standardized commenting form which included the lines number of the before mentioned dissemination materials and the category of the comment (major/minor/linguistic) for each comment. After closing this consultation period, the working group developed consolidated answers for each comment. The working paper and the framework itself were amended if deemed necessary.

Implementation

Broad internal discussions were performed to integrate the results of pilot (round #1) and retrospective testing (round #2–4) as well as the input from stakeholders. These discussions were used to elicit questions to be answered in order to reach alignment on procedural and methodological details for the implementation of concluding on CAB (i.e., integration into the Department's assessment procedure). The Department's internal knowledge repository was used to capture all relevant findings and to facilitate dissemination among assessors. Eventually, a summary of the framework was also anticipated for publication to inform stakeholders.

Results

Review and Retrospective Analysis of Reimbursement Dossiers

As a result of the review, several factors were identified which can influence the conclusion on CAB in the assessments of HTA bodies: disease severity, innovation, unmet need, relevance of endpoints, magnitude of benefit, quality of evidence, public health importance, and patient value. Table 1 presents the distributions of CAB/CAV categories of dossiers submitted to NIPN in 2019–2020 as classified by HAS and published CAV/CAB categories of dossiers from HAS and IQWiG in 2011–2017 (8). The distribution of CAV categories among dossiers evaluated by both HAS and NIPN in 2019–2020 does not differ strikingly from those evaluated by HAS in 2011–2017 (8). However, there was a substantial proportion of dossiers which was not evaluated by HAS but was submitted to NIPN (23 percent). Moreover, between systems used by different countries the differences are more prominent: IQWiG tends to consider more pharmaceutical products to be characterized with "Major" clinical benefit, whereas HAS rarely uses this category (8). These results together emphasized the need for developing a system tailored to Hungary.

Most evaluated compounds belonged to the field of oncology and hematology (41 percent), followed by rheumatology (11 percent), neurology and psychiatry (9 percent), cardiovascular diseases (9 percent), and immunology (5 percent).

Internal Development

Based on our targeted review of other HTA bodies' guidance, we concluded that the determination of CAB is influenced by the nature of the considered endpoint(s), the extent of the difference compared to the comparator, and the reliability and accuracy of the results.

To align the current practices with the formulation of the conclusion on the CAB, the four domains in the developed framework are considered as equal contributors to the conclusion:

(i) information on the relevance of the considered clinical endpoints;

(ii) the existence/extent of the added benefit, and

(iii)–(iv) the quality of evidence supporting it, which has two subdomains: (iii) the level of evidence and (iv) the risk of bias (RoB) associated with it in the cases of clinical trials.

To evaluate the relevance of an endpoint, we rely on the guidelines's recommendations (11–13): In the case of a life-threatening

https://doi.org/10.1017/S0266462322000411 Published online by Cambridge University Press

Table 2. Concordance Between Assessors During the Three Rounds of Retrospective Assessments

Pilot test round no.	Endpoint relevance	Extent of CAB			Level of evidence	Risk of bias
		Category	Baseline score with ESMO-MCBS	Adjusted score with ESMO-MCBS		
Round #2 $n = 7$	85.7%	83.3%	100.0%	76.2%	85.7%	57.1%
Round #3 $n = 9$	77.8%	76.9%	100.0%	77.8%	88.9%	85.2%
Round #4 $n = 7$	61.9%	90.5%	76.2%	76.2%	90.5%	90.5%
Average (SD)	75.1% (15.1)	83.5% (6.2)	92.1% (11.9)	76.7% (13.8)	88.4% (2.7)	77.6% (13.6)

Notes. Round #1 was a small volume pilot therefore that is not presented here. In each round, three different dossiers were retrospectively evaluated. Each round consisted of three previously not evaluated dossiers. From round to round, the complexity of the evaluated dossiers increased.
CAB, clinical added benefit; ESMO-MCBS, European Society for Medical Oncology – Magnitude of Clinical Benefit Scale; *n*, number of participating medical assessors.

disease, mortality as an endpoint or survival endpoints are the most relevant and morbidity and/or quality of life are secondary. In case of non-life-threatening diseases, morbidity and quality of life endpoints are preferred. We consider endpoints relevant for patients if they are associated with either improved overall survival or improved/sustained quality of life.

We decided to use the ESMO-MCBS for scoring the extent (magnitude) of clinical benefit of antineoplastic drugs. We assigned categories for determining the extent of benefit from the ESMO-MCBS scores (Supplementary Table 1). Scores A and 5 were considered as "Major added benefit," scores B, 4, and 3 as "Important," and scores C and 2–1 were categorized as "Minor added benefit." In cases where statistically significant difference on a relevant endpoint in the PICO of the reimbursement submission was not observed (e.g., because a single-arm study does not have the comparator determined by PICO), we assigned the categories 'No proof of benefit' or the category 'Not quantifiable' in cases where methodological issues emerged as well.

In the case of indirect comparisons, where the dual rule of the ESMO-MCBS cannot be used (only relative efficacy can be derived, direct comparison is not available), we are not able to determine the *extent* of CAB. However, the conclusion on the existence of CAB always preludes the conclusion on its extent. First our conclusion includes information on the *existence* of the CAB: (i) the *existence* of the CAB is possible; (ii) it is not proven (=No proof of benefit), or (iii) it cannot be determined based on the presented evidence (=Not quantifiable). If the *existence* of CAB is possible—in the cases of direct comparisons—we can extend the conclusion with determining its *extent*.

Classification of the level of evidence was adapted from one of the published SOPs of ESMO (14). Different levels of evidence were merged into a simplified rating scale with the categories of high–moderate–low levels of evidence (Supplementary Table 3).

- HIGH: large, good quality randomized controlled trials (RCT) and meta-analyses of these without considerable heterogeneity.
- MODERATE: small RCTs or large RCTs with susceptible bias and meta-analyses of these or meta-analyses with considerable heterogeneity or indirect comparison.
- LOW: cohort studies, case reports, and indirect comparisons (15) where the methodology is not clearly presented in the submitted application for reimbursement or if the indirect comparison carries serious methodological flaws.

Due to capacity constraints, the conclusion refers to an external source of RoB assessment (e.g., from IQWiG/EUnetHTA reports, Cochrane, or other published sources) for the time being.

Retrospective Testing

The framework was tested in the field of oncology. In the pilot round #1, the results were consistent with the scores published on the ESMO website.

As a next step, all clinical assessors of the Department retrospectively evaluated submission dossiers in three consecutive rounds (three different dossiers in each round, with increasing complexity). Average results had more than 75 percent concordance between assessors on each element of the conclusion in all rounds, except for two categories ("RoB" (51.7 percent in round #2) and the category "Endpoint relevance" (61.9 percent in round #4; Supplementary Table 4). The highest concordance rates between assessors were observed regarding the levels of evidence and the baseline ESMO-MCBS scores (Table 2). The causes of the discordance, in the cases of adjusted ESMO-MCBS scores, were due to the different evaluation of toxicity, or in cases of dossiers requiring adjustments based on progression-free survival (PFS) plateaus. As for the categories representing the extent of benefit, the concordance between assessors was high in most cases. The low concordance initially observed regarding the RoB was caused by the differences on the whole study level versus the RoB associated with the relevant endpoints. The lowest concordance was found regarding the endpoint relevance.

Each round was followed by detailed discussions, and the internal guidance was consensually revised to cover the questions raised during assessments. These questions concerned incomplete data regarding toxicities, evaluation of RoB, and decisions on endpoint relevance. The experts of ESMO have been consulted via email on certain issues related to subgroup analyses, evaluation of PFS plateaus, and indirect comparisons. For clarifying questions regarding toxicity-related downgrading, the Department of Pharmacovigilance at NIPN was contacted.

Feedback from Stakeholders

Finally, ten stakeholders responded to the call from which two responded only to the questionnaire and two responded only to the commenting form, while six participants provided feedback via both instruments. We received one response from academic centers and one from public bodies and two responses from each of the following entities: patient organizations, medical societies, industry associations, and consultancy firms.

A total of seventy-two comments were received from eight participants on the standardized commenting form. In general, the initiative to develop a new procedure to assess CAB was welcomed by all stakeholders and it was also agreed that the

https://doi.org/10.1017/S0266462322000411 Published online by Cambridge University Press

proposed framework would greatly contribute to an increase in the quality of assessment reports in Hungary. Most of the responses were concerned with the following topics: (i) concepts used in the framework; (ii) aspects considered during the formulation of the conclusion on CAB, and (iii) methodology to decide on the extent of CAB. Patient organizations and medical associations shared their opinion mainly about relevant outcomes, emphasizing that patients' perspective is important to be included. Industry associations, consultancy firms, and academic centers provided in-depth feedback concerning methodological issues, including the assessment of available evidence, the relevance of endpoints, and the classification of CAB in the final conclusion. After careful consideration, their suggestions were implemented and the framework was refined accordingly.

Eight out of ten stakeholders filled in the questionnaire. Answers were regrouped as concordant (fully agree, rather agree) and discordant (fully disagree, rather disagree) responses. The system describing the quality of clinical evidence was welcomed. Opinions were also generally supportive about the scale used to score the extent of CAB and all of the respondents had a positive attitude toward the implementation of the ESMO-MCBS and shared the opinion that the proposed framework might improve the quality of HTA reports (questions D2 and A4, respectively, in Figure 2).

Kendall's Coefficient of Concordance (Kendall's W) was calculated to assess the concordance of responses between stakeholders. The coefficient was found to be significant with a value of .367 ($p = .002$) that can be interpreted as a fair agreement in this context (16).

Implementation

In terms of operative issues, the final version the working paper was shared with the stakeholders and the rollout date of 1 Jan 2022 was agreed upon internally; a set of common phrases for the assessment template were drafted to ease reporting. A sequential escalation procedure was designed to support the clinical assessors, should uncertainty in formulating the conclusion on CAB arise.

Figure 3 presents the procedural steps of the consolidated framework as it is formally implemented in the assessment procedure. Determination of CAB is based on the scientific evidence submitted by the Applicant, determining the PICO structure of the analysis. The targeted literature review serves to decide whether higher quality scientific evidence is available.

The HTA report examines and uses common phrases to consistently report the relationship between the characteristics of the CAB and the incremental health gain quantified in the cost-effectiveness analysis. The outcomes considered in determining the CAB are expected to be the same or overlap to a large extent with the outcomes that are the source of the incremental health gain.

Discussion

This paper describes the development, pilot and retrospective testing, stakeholder input assessment, and implementation of a framework designed for concluding on CAB in Hungary. Although many elements of this framework are already in use and assessed in the critical appraisals, a standardized and transparent system is lacking.

Using frameworks similar to the one presented here to guide the value assessment of health technologies is not unique. Recent research shows that such value frameworks are being tailored to geographic regions and types of health technologies (17). Using existing frameworks, like ESMO-MCBS, in a national setting is not unique either. In Korea, the American Society of Clinical Oncology and the ESMO-MCBS were adapted to produce a reliable framework (18). In Canada, multicriteria decision analysis methods were applied to the development of a value assessment framework for antineoplastic drugs. In addition, researchers validated the

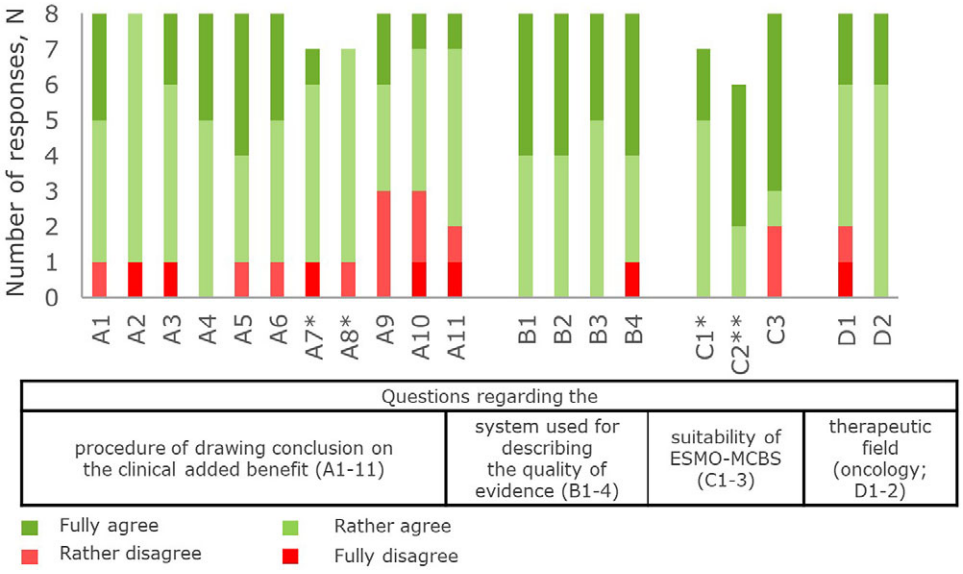

Figure 2. Distribution of stakeholder responses.
In total, ten stakeholders provides feedback from which two did not answer the questionnaire and two did not give any additional comments. Figure shows the responses of eight stakeholders to the questionnaire.
*missing values = 1; **missing values = 2. A1-D2: IDs of questions of the stakeholder questionnaire (see Supplementary: stakeholder questionnaire). Domain A: feedback regarding the procedure of drawing conclusion on CAB (clinical added benefit). Domain B: feedback regarding the system used for describing the quality of evidence. Domain C: feedback regarding the scale used for scoring the extent of CAB. Domain D: feedback regarding the therapeutic field chosen for introduction (solid tumors).

https://doi.org/10.1017/S0266462322000411 Published online by Cambridge University Press

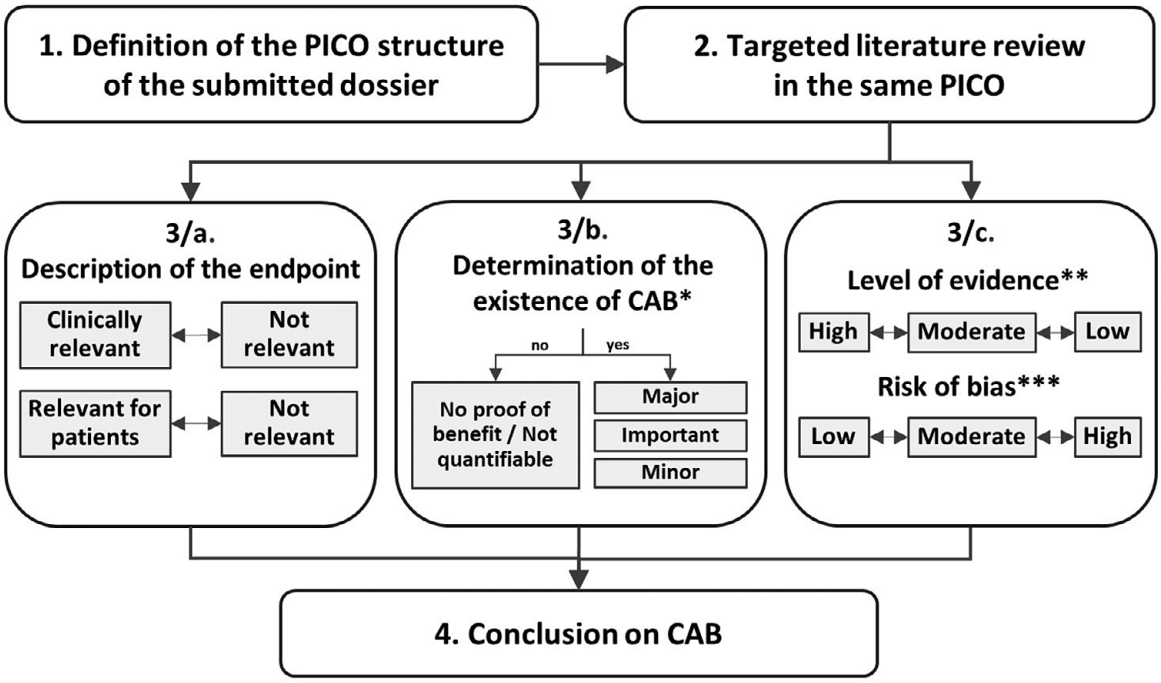

Figure 3. Schematic presentation of the process of formulating conclusion on CAB.
*In cases where the conclusion cannot be drawn on the extent, we provide conclusion only on the existence of CAB.
**High: large, good quality randomized controlled trials (RCT) and meta-analyses of these without considerable heterogeneity. Moderate: small RCTs or large RCTs with susceptible of bias and meta-analyses of these or meta-analyses with considerable heterogeneity or indirect comparisons. Low: cohort studies, case reports, and indirect comparisons where the methodology is not clearly presented in the submitted application for reimbursement or if the indirect comparison carries serious methodological flaws. Further details are presented in Supplementary Table 2.
***Categories based on use of an internationally accepted tool (e.g., GRADE or Cochrane RoB2).
CAB, clinical added benefit; PICO, population-intervention-comparator-outcome.

framework by assessing the correlation of the resulting scores and ESMO-MCBS thresholds for meaningful benefit (19). In Slovenia, the overall time to access novel antineoplastic pharmaceuticals and its correlation with ESMO-MCBS scores were assessed. Researchers found that time to access is similar for drugs with or without substantial CAB. According to their conclusion, integrating the ESMO-MCBS into reimbursement deliberations could improve access to drugs with substantial clinical benefit (20). A similar analysis is planned in Hungary after our framework is routinely used.

The proposed procedure could serve as a pillar for the characterization of the relationship between the CAB and the incremental health gain quantified in the cost-effectiveness analysis. A recently published article about HAS's system (21) found convergence between the independent clinical and economic appraisal processes: the CAV rating is positively associated with the disease severity, the quality-adjusted life-year gain provided by the drug, and the validation of the incremental cost-utility ratio in the Economic Opinion. A similar validation process could be performed in the case of our framework in the future.

One limitation of this research is that the extent of CAB is currently based on the ESMO-MCBS. Although 41 percent of reimbursement dossiers concern treatments for oncological diseases, there is a need to develop guidance for other therapeutic areas. In cases where the ESMO-MCBS cannot be used (indirect comparisons and therapeutic areas other than solid tumors) we can only give conclusions on the possibility of the *existence* of CAB instead of categorizing its *extent.* This is an area for further development.

We are aware that the ESMO-MCBS is not free from bias itself (22). The developing authors proposed an updated version of the scale, which could deal with hematological malignancies as well and incorporates guidelines for meta-analyses to be evaluated (23). It is also notable that ESMO states that therapies with scores A–B or 4–5 are characterized with substantial clinical benefit. Our categorization is somewhat more rigorous; we consider therapies scoring A or 5 as representing a major added benefit. We argue that innovative therapies often come at a significant cost; it is important to distinguish therapies with the highest CAB. Adaptation of the ESMO-MCBS for HTA purposes is not exceptional; the Austrian Institute for Health Technology Assessment also incorporated it in its assessments, and they also adapted the scoring system for their purposes (24). The authors found that a minority of the assessed dossiers met the meaningful benefit criteria and concluded that both the original and the modified ESMO-MCBS can help to identify potentially beneficial cancer medications and thus support the fair allocation of limited healthcare resources. Furthermore, a recent analysis of oncological approvals in Europe highlighted that using ESMO-MCBS in the assessment of drug dossiers and reimbursement negotiations, especially for drugs with low or questionable clinical benefit, might be beneficial (25).

One might notice that the last round of retrospective assessments resulted in a slightly lower concordance rate than the previous ones in the cases of endpoint relevance and adjusted ESMO-MCBS scores. In our opinion, the reason for this was the increasing complexity of dossiers selected for retrospective analysis (more advanced diseases often with surrogate endpoints and older dossiers where updated OS results were available). The adjustments

https://doi.org/10.1017/S0266462322000411 Published online by Cambridge University Press

for ESMO-MCBS baseline scores can result a higher value if reduced toxicity or improvement in quality of life is observed. In studies where the primary endpoint is not OS, adjustments consider the survival advantage as well. In the cases of adjusted ESMO-MCBS scores in our retrospective assessments one cause of discordance was the evaluation of toxicity, since formal statistical hypothesis testing was rarely carried out. Determination of PFS plateaus is also a critical point and requires aid from a statistical expert. The relevance of endpoints might be up for scientific debate. According to the EUnetHTA and EMA guidelines, the primary interest is survival in life-threatening diseases, but in individual situations, other survival-related outcomes could serve as a good primary endpoint if the OS results are supporting it (e.g., diseases where progression is a validated surrogate for OS). These situations are not always self-evident and require a consolidated position of the assessors.

The lack of validation of the proposed framework to already existing value frameworks may be viewed as a limitation. However, we argue that a perfect alignment could also mean that it is unnecessary to use an assessment method other than the reference. One may also suggest that the CAB assessment should be extended beyond the scientific evidence directly supporting the economic analysis through conducting a full-scale systematic review. We reply that we follow the concept of critical assessment ensuring the best possible evidence is used for the analysis by a necessary confirmatory, targeted literature review. Using the assessment of an external party on RoB can be identified as a limitation, yet we consider it to be a necessary temporary solution to optimize resource use.

Our conclusion on CAB consists of four elements, which are intentionally not merged into a single score. This way the potential loss of relevant information is lower; however, direct comparison with other European systems is more difficult.

According to their feedback, the concept of the framework is welcomed by local stakeholders, who generally supported the framework and its introduction into the local HTA process. However, proposals concerning some methodological issues were received and refinements of the framework were amended as a response to stakeholder feedback. In the questionnaire, most disagreements were found regarding the framework's usefulness in the learning process of stakeholders and details of the ESMO-MCBS, primarily due to the limited awareness of stakeholders about these attributes. It should also be mentioned as a limitation that further assessment of stakeholder answers was restricted due to the low number of responses.

The public consultation of the procedure contributes to the transparency of the development and strengthens the validity of its use for decision making. The conclusion on CAB can accompany the formerly proposed conceptual framework (26) of assessing sources of uncertainty in the cost-effectiveness analyses submitted in the reimbursement dossiers. Finally, the assessment of CAB may contribute to transparently formulate the position of NIPN in the appraisal committees.

Some aforementioned methodological issues are subjects of future research. First, additional questions may arise from the definition of a clinically or patient-relevant endpoint that can be well defined for patients with certain diseases, for other diseases might be the subject of debate or evolving over time with the development of medical science. First, our basic rule is that a CAB measured on a patient-relevant endpoint must be supported with either increased life expectancy or increased/sustained quality of life. Second, the conclusion on CAB does not currently include information on the extent and significance of the innovation, unmet medical need, the severity of the disease, and its public health significance, all of which may be important considerations for deciding on reimbursement.

Conclusion

We demonstrated the development, retrospective testing, feedback assessment, and implementation of a procedure for describing the CAB of pharmaceuticals, specifically tailored for the local health policy environment in Hungary.

Considering its strengths and limitations, we believe the proposed procedure is suitable for routine use in the local decision making on pharmaceutical treatment of solid tumors in Hungary. However, extensions of the framework are required to cover more disease areas to provide structured and comparable conclusions on CAB of innovative health technologies.

Supplementary Materials. To view supplementary material for this article, please visit http://doi.org/10.1017/S0266462322000411.

Conflict of interest. The authors reported no conflicts of interest.

Funding. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

References

1. **Drummond MF**, **Schwartz JS**, **Jönsson B**, et al. (2008) Key principles for the improved conduct of health technology assessments for resource allocation decisions. *Int J Technol Assess Health Care* **24**, 244–258; discussion 362–368.
2. **Akehurst RL**, **Abadie E**, **Renaudin N**, **Sarkozy F** (2017) Variation in health technology assessment and reimbursement processes in Europe. *Value Health* **20**, 67–76.
3. **Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen**. *General Methods v6.1, 2022* [Internet] [accessed 8 Feb 2022]. Available at: https://www.iqwig.de/en/methods/methods-paper.3020.html.
4. **Haute Autorité de santé**. *Transparency Committee Doctrine - Principles of medicinal products assessment and appraisal for reimbursement purposes* [Internet] [accessed 8 Feb 2022]. Available at: https://www.has-sante.fr/upload/docs/application/pdf/2019-07/doctrine_de_la_commission_de_la_transparence_-_version_anglaise.pdf.
5. **Kaló Z**, **Gheorghe A**, **Huic M**, **Csanádi M**, **Kristensen FB** (2016) HTA implementation roadmap in central and Eastern European Countries. *Health Econ* **25**(Suppl 1), 179–192.
6. **Gulácsi L**, **Rotar AM**, **Niewada M**, et al. (2014) Health technology assessment in Poland, the Czech Republic, Hungary, Romania and Bulgaria. *Eur J Health Econ* **15**(Suppl 1), S13–S25.
7. **Schünemann HJ**, **Wiercioch W**, **Etxeandia I**, et al. (2014) Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ* **186**, E123–E142.
8. **Boucaud-Maitre D**, **Berdaï D**, **Salvo F** (2021) Added therapeutic value of medicinal products for French and German health technology assessment organizations: a systematic comparison. *Value Health* **24**, 346–352.
9. **EUR-Lex**. *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on health technology assessment and amending Directive 2011/24/EU. COM/2018/051 final - 2018/018 (COD)* [Internet] [accessed 7 Jan 2022]. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32021R2282.
10. **Cherny NI**, **Sullivan R**, **Dafni U**, et al. (2015) A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS). *Ann Oncol* **26**, 1547–1573. Erratum in *Ann Oncol.* 2017;28:2901–290

11. **EUnetHTA**. *Guideline - Endpoints used for relative effectiveness assessment: clinical endpoints* [Internet] [accessed 3 Oct 2021]. Available at: https://www.eunethta.eu/wp-content/uploads/2018/02/WP7-SG3-GL-clin_endpoints_amend2015.pdf.
12. **EUnetHTA**. *Guideline - Endpoints used in relative effectiveness assessment: surrogate endpoints* [Internet] [accessed 3 Oct 2021]. Available at: https://www.eunethta.eu/wp-content/uploads/2018/03/surrogate_endpoints.pdf.
13. **EMA**. *Clinical efficacy and safety guidelines* [Internet] [accessed 3 Oct 2021]. Available at: https://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines/clinical-efficacy-safety-guidelines.
14. **ESMO**. *ESMO Standard Operating Procedures (SOPs) for Clinical Practice Guidelines (CPGs) and ESMO Magnitude of Clinical Benefit (ESMO-MCBS) scores* [Internet] [accessed 22 Nov 2021]. Available at: https://www.esmo.org/content/download/77789/1426712/file/ESMO-Clinical-Practice-Guidelines-Standard-Operating-Procedures.pdf.
15. **Indirectness of evidence**. In: Schünemann H, Brożek J, Guyatt G, Oxman A, editors, *GRADE handbook for grading quality of evidence and strength of recommendations.* Published online October 2013. [accessed 22 Nov 2021]. Available at: https://gdt.gradepro.org/app/handbook/handbook.html.
16. **Gisev N**, **Bell JS**, **Chen TF** (2013) Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Res Social Adm Pharm* **9**, 330–338.
17. **Augustovski F**, **Alfie V**, **Alcaraz A**, et al. (2021) A Value Framework for the Assessment of Diagnostic Technologies: A Proposal Based on a Targeted Systematic Review and a Multistakeholder Deliberative Process in Latin America. *Value Health* **24**, 486–496.
18. **Bae G**, **Bae S**, **Lee D**, et al. (2021) Value frameworks: adaptation of Korean versions of value frameworks for oncology. *Int J Environ Res Public Health* **18**, 3139.
19. **Ezeife DA**, **Dionne F**, **Fares AF**, et al. (2020) Value assessment of oncology drugs using a weighted criterion-based approach. *Cancer* **126**, 1530–1540.
20. **Janzic U**, **Knez L**, **Janzic A**, **Cufer T** (2019) Time to access to novel anticancer drugs and the correlation with ESMO-Magnitude of Clinical Benefit Scale in Slovenia. *Expert Rev Pharmacoecon Outcomes Res* **19**, 717–723.
21. **Kergall P**, **Autin E**, **Guillon M**, **Clément V** (2021) Coverage and pricing recommendations of the French National Health Authority for innovative drugs: a retrospective analysis from 2014 to 2020. *Value Health* **24**, 1784–1791.
22. **Gyawali B**, **de Vries EGE**, **Dafni U**, et al. (2021) Biases in study design, implementation, and data analysis that distort the appraisal of clinical benefit and ESMO-Magnitude of Clinical Benefit Scale (ESMO-MCBS) scoring. *ESMO Open* **6**, 100117.
23. **Knapen DG**, **Cherny NI**, **Zygoura P**, et al. (2020) Lessons learnt from scoring adjuvant colon cancer trials and meta-analyses using the ESMO-magnitude of clinical benefit scale V.1.1. *ESMO Open* **5**, e000681.
24. **Grössmann N**, **Wolf S**, **Rothschedl E**, **Wild C** (2021) Twelve years of European cancer drug approval—A systematic investigation of the 'magnitude of clinical benefit'. *ESMO Open* **6**, 100166.
25. **Falcone R**, **Lombardi P**, **Filetti M**, et al. (2022) Oncologic drugs approval in Europe for solid tumors: overview of the last 6 years. *Cancers (Basel)* **14**(4), 889.
26. **Merész G**, **Dóczy V**, **Hölgyesi Á**, **Németh G** (2021) [A practical demonstration of critically assessing health economic analyses]. Az egészséggazdaságtani elemzések kritikai értékelési gyakorlatának bemutatása. *Lege Artis Medicinae* **31**, 519–529. Hungarian.

https://doi.org/10.1017/S0266462322000411 Published online by Cambridge University Press