EXPERT REVIEW OF PHARMACOECONOMICS & OUTCOMES RESEARCH
https://doi.org/10.1080/14737167.2019.1559732

Taylor & Francis
Taylor & Francis Group

ORIGINAL RESEARCH

Is the National Institute for Health and Care Excellence (NICE) in England more 'innovation-friendly' than the Federal Joint Committee (G-BA) in Germany?

Ramon Schaefer[a,b,c] and Michael Schlander[a,b,c]

[a]Division of Health Economics, German Cancer Research Center (Deutsches Krebsforschungszentrum, DKFZ), Heidelberg, Germany; [b]Mannheim Medical Faculty, University of Heidelberg, Mannheim, Germany; [c]Institute for Innovation & Valuation in Health Care (InnoVal[HC]), Wiesbaden, Germany

ABSTRACT

Objectives: Our study explores whether, and how, different methodological choices are associated with different health technology assessment (HTA) outcomes. We focus on the Federal Joint Committee (Gemeinsamer Bundesausschuss, G-BA) in Germany and the National Institute for Health and Care Excellence (NICE) in England. Both agencies may be considered as exemplars for the application of the principles of evidence-based medicine and the logic of cost-effectiveness, respectively.

Methods: We extracted data from all publically available G-BA appraisals until April 2015, as well as all NICE single technology appraisals completed during this period. We compared HTA results for matched condition-intervention pairs by G-BA and NICE, and explored other factors including therapeutic area, clinical effectiveness and cost-effectiveness.

Results: NICE issued guidance for 88 technologies (125 subgroups) and recommended 67/88 technologies (99/125 subgroups). G-BA completed 105 appraisals (226 subgroups) and determined additional benefit for 64/105 appraisals (90/226 subgroups). We identified 37 matched pairs; for 24/37 drugs, evaluations diverged. NICE recommended 78% (29/37) of technologies appraised, whereas G-BA confirmed additional benefit for 57% (21/37) only ($p < 0.05$).

Conclusions: NICE evaluates new drugs more favorably than G-BA. However, our analysis suggests differences by therapeutic area. Results indicate that different methods are associated with systematic differences in HTA outcomes.

ARTICLE HISTORY
Received 2 October 2018
Accepted 13 December 2018

KEYWORDS
Condition-intervention pair; early benefit assessment; evaluation; health technology assessment; single technology appraisal

1. Introduction

Concerns about escalating health care spending date back to the 1970s and have since been a strong motive behind the increasing use of health economic evaluations. Similarly, the introduction of systematic health technology assessments (HTAs) by the United States Office of Technology Assessment (OTA) in 1975 can be traced back to the desire for effective and efficient use of costly health technologies [1,2]. Today, HTA is understood as a multidisciplinary process supporting decision-making in health care, based on scientific and non-scientific evidence [1–3]. Although HTA has been described as a comprehensive evaluation method, in practice it rests predominantly on two pillars: the assessment of clinical benefit drawing on principles of evidence-based medicine (EBM), and an evaluation of efficiency, usually by means of a variant of cost-effectiveness analysis (CEA) [2]. Unsurprisingly, given the objective to contain health care spending and to provide access to innovative and at the same time affordable treatments, the creation of official HTA agencies has been accompanied by controversial debate and outright concerns about 'rationing' [4].

This was also true for the National Institute for Health and Care Excellence (NICE), which was established as a Special Health Authority within the United Kingdom (UK) National Health Service (NHS) in 1999 [5]. Its first technology appraisal resulted in the rejection of zanamivir – a drug developed by the British pharmaceutical giant Glaxo – for the treatment of influenza in October 1999. After the drug was resubmitted by the manufacturer, however, NICE revised its decision for at least one subgroup in November 2000. Finally, zanamivir has been recommended for the treatment of high-risk individuals and made available within the NHS [6].

Since its inception, NICE has rapidly gained a reputation as an international role model for HTAs including cost utility analysis (CUA), combining high methodological standards and transparency of assessments with the use of somewhat broader criteria in appraisals [5,7,8]. Issued by NICE´s Centre for Health Technology Evaluation, technology appraisals are recommendations on the use of existing and new health technologies, which are mandatory for the NHS in England and Wales. The evaluation process

CONTACT Ramon Schaefer ramon.schaefer@dkfz.de German Cancer Research Center (Deutsches Krebsforschungszentrum, DKFZ), Im Neuenheimer Feld 280, 69120, Heidelberg, Germany

Parts of the present materials and data were presented at the ISPOR *(International Society for Pharmacoeconomics & Outcomes Research)* Annual European Conferences in Vienna/Austria, October 29 – November 02, 2016, and in Glasgow/Scotland, November 04 – 08, 2017, as well as at the HTAi *(Health Technology Assessment International)* Annual Meeting in Rome/Italy, June 17 – 21, 2017.

Supplemental data for this article can be accessed here.

for new technologies consists of three main stages: scoping, assessment, and appraisal (and an optional fourth stage: appeal). The evaluation process can take the form of a multiple technology appraisal (MTA) that results in guidance for single or multiple products, devices or other technologies, with one or more indications. An alternative process, single technology appraisals (STAs), was developed for the rapid review of new technologies for a single indication; it takes approximately 35 weeks from initiation of the appraisal to guidance publication [5]. For new pharmaceutical products, NICE primarily relies on incremental cost-effectiveness ratios (ICERs) as an indicator of 'value for money' [9]. Health technologies with an ICER below £20,000 per quality-adjusted life year (QALY) gained are most likely to be recommended by NICE. Likewise, most technologies with ICERs in the range of £20,000–£30,000 are evaluated positively by NICE, whereas newly authorized drugs with ICERs above the upper benchmark are usually rejected by NICE [10–12]. Exceptions were introduced for end of life (EoL) treatments (primarily cancer drugs) and by establishing a separate highly specialised technology (HST) evaluation program (for ultra-orphan indications). Cancer drug appraisals have remained particularly controversial against the background of an apparent increase of negative evaluation outcomes [13].

A different evaluation approach was adopted by the Institute for Quality and Efficiency in Health Care (IQWiG) and the Federal Joint Committee (Gemeinsamer Bundesausschuss, G-BA) in Germany. Both institutions rejected the incremental cost per QALY metric as a measure of 'value for money'. Instead, official HTAs in Germany rest on an assessment of comparative effectiveness based on stringent application of the principles of EBM [14,15].

With the enactment of the Pharmaceutical Market Restructuring Act (Arzneimittelmarktneuordnungsgesetz, AMNOG) in 2011, Germany set up a two-stage process for early benefit assessment (EBA) of newly authorized pharmaceuticals [16,17]. According to §35a Social Code Book V, G-BA appraisals of additional clinical benefit will inform subsequent decisions on pricing and reimbursement within the insurance-based health system. Manufacturers are required to submit a comprehensive clinical value dossier at the time of launch. Usually, G-BA commissions IQWiG to conduct a formal review of the dossier, which provides the basis for deliberation and final decision by the G-BA on the extent of (and level of confidence in) an added clinical benefit. If the G-BA accepts a manufacturer's claim of additional benefit, the National Association of Statutory Health Insurance Funds (Spitzenverband Bund der Krankenkassen, GKV-SV) and the pharmaceutical manufacturer will negotiate a reimbursement price. If they do not reach an agreement, an arbitration board will determine the reimbursement price. However, if and when G-BA does not confirm the existence of additional benefit, the new drug will be set to become part of the reference price system [17]. Health economic evaluations may be initiated only after an arbitration price was set, and

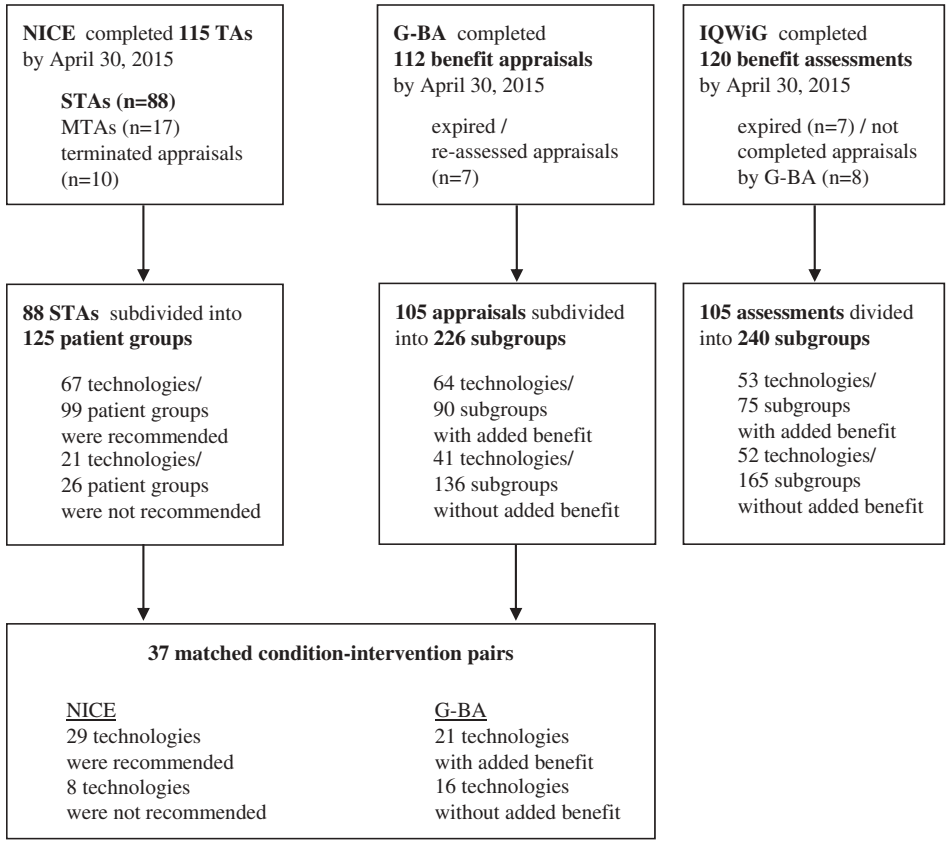

Figure 1. Overview on NICE STA guidance, G-BA appraisals, IQWiG dossier assessments and matched condition-intervention pairs. G-BA, Gemeinsamer Bundesausschuss; IQWiG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; MTA, multiple technology appraisal; NICE, National Institute for Health and Care Excellence; STA, single technology appraisal; TA, technology appraisal.

would require the request of either manufacturer or sick funds. Hence, formal economic analyses have not been used in the context of EBAs or subsequent pricing and reimbursement decisions in Germany.

Like the approach in England, the choice of German policy-makers to rely heavily on principles of EBM has proved workable. Yet, there remain areas of controversy, such as the reluctance by G-BA and IQWiG to accept data showing effects on surrogate endpoints only [16,18]. This may imply higher hurdles for drugs for chronic diseases compared to an approach that allows modeling of future events based on surrogate endpoints. On the other hand, legal provisions in Germany specify that drugs with an orphan designation (use of which must not result in statutory health insurance (SHI) acquisition costs exceeding €50 million per annum) by definition are assumed to confer some additional benefit. Accordingly, IQWiG does not asses orphan drugs, and G-BA will only evaluate the extent of the additional benefit, which in most cases has been determined as either non-quantifiable or minor [19]. In Germany, manufacturers may decide to 'opt-out', i.e. to discontinue distribution of a new product within four weeks after G-BA appraisal. By the end of 2015, manufacturers had withdrawn 20 new drugs from the German market, either because of negative appraisals by G-BA or due to low reimbursement prices [20]. These observations contributed to a widely held perception that G-BA decisions may be more restrictive – and hence less 'innovation-friendly' – than NICE appraisals [16,21].

Against this background, the present study was designed to assess the hypothesis that different methodological choices made by NICE and by G-BA/IQWiG are indeed associated with different HTA outcomes, and to explore any patterns emerging from the observed differences (if any). A comparative study of HTA outcomes by NICE and G-BA/IQWiG should be of interest to international health care policy-makers, because both government independent official HTA agencies have been broadly recognized as exemplars for the implementation of the logic of cost effectiveness and the principles of EBM, respectively.

We will first provide a descriptive report of the overall outcomes of STAs by NICE and of EBAs by G-BA/IQWiG during the observation period of the study, i.e. between January 2011 and April 2015. Secondly, we check for both agencies whether their respective overall HTA outcomes result from a consistent application of their officially stated evaluation criteria. In order to avoid comparing apples and oranges, we then identify matched condition-intervention pairs, and assess, whether our data confirm the prevailing perception that outcomes of NICE STAs differ from G-BA/IQWiG EBAs.

2. Methods

2.1. Dataset

Both NICE and G-BA/IQWiG publish detailed information about ongoing and completed technology assessments and appraisals on their respective websites [22,23]. We first identified all NICE STAs [22] and all G-BA EBAs [23] completed between 1 January 2011 and 30 April 2015 *(Figure 1)*.

We then extracted data from guidance issued by NICE and, in case of missing information, from Evidence Review Group (ERG) reports, representing an independent review of the manufacturer´s dossier and clinical advice. Data used for subsequent analyses included therapeutic area and benefit assessment results (recommended/not recommended), size of patient population (eligible for treatment) and subpopulation (patient groups), clinical evidence (existence of relevant randomized controlled trials, RCTs, in the manufacturer submission), cost-effectiveness (ICER per QALY based on the Appraisal Committee´s determination), annual drug acquisition costs per patient, and EoL criteria (if appropriate).

From published G-BA appraisals (and in case of missing data, completed IQWiG dossier assessments), we extracted benefit determination outcomes by assessment category (in terms of certainty – proof, indication, hint, and extent – major, considerable, minor, non-quantifiable, no added benefit, lesser benefit), therapeutic area, size of patient population and subpopulation (subgroups), clinical evidence (in terms of the use of relevant RCTs in manufacturer dossiers), patient-relevant endpoints (focusing on mortality, morbidity and health-related quality of life, HRQoL), as well as annual treatment costs per patient.

2.2. Statistical analysis of HTA outcomes by G-BA/IQWiG and NICE

Based on previous study results, we determined drivers for statistical analysis that appeared most likely to drive assessments and appraisals by NICE [10–12] and G-BA/IQWiG [14,24] the most. For NICE we analyzed the ICER per QALY gained (implicit ICER thresholds used as decision criterion; <£20,000/ £20,000–£30,000/ >£30,000/ N/A, not applicable) [10–12,25]. In addition, the impact of clinical evidence (existence of relevant RCTs in the manufacturer submission; yes/no) as well as the application of EoL criteria (if appropriate; yes/no) for the assessment of new drugs with an ICER above £30,000 (primarily focusing on cancer drug assessments). By definition, patient relevant-endpoints are the most relevant criteria for IQWiG dossier assessments and G-BA appraisals [14,24]. We thus focused on the impact of patient-relevant outcomes (significant superiority of either mortality, morbidity or HRQoL compared with the appropriate comparative therapy, ACT; yes/no) as well as the impact of clinical evidence (relevant RCTs submitted by manufacturer; yes/no) for both G-BA appraisals and IQWiG dossier assessments.

We then performed a multivariate linear analysis of dossier assessments by NICE as well as G-BA/IQWiG, respectively. Because orphan drugs in Germany are assumed to confer additional benefit, we excluded these drugs for statistical analysis of G-BA appraisals and IQWiG dossier assessments. We tested the impact of HTA results and relevant variables (i.e. previously specified evaluation criteria of both agencies) for statistical significance using the chi-square test or Fisher-Yates test.

2.3. Comparative analyses

We compared HTA outcomes (including orphan drugs) by NICE and G-BA/IQWiG at two levels: overall observations (total sample retrieved for analysis), and matched condition-intervention pairs.

Table 1. Matched condition-intervention pairs: NICE STA guidance and G-BA appraisals.

		G-BA appraisals				NICE STAs		
		Decision:	Patient-relevant endpoints[a]:			Guidance:	ICER per QALY	EoL criteria
Health technology	Therapeutic area	Added benefit	Mortality	Morbidity	HRQoL	Recommended	*(in GBP)*	*(if applicable)*
Retigabine	Neurological	No				Yes	40,000–60,000	
Ticagrelor	Cardiovascular	Yes	+	+		Yes	7,897	
Apixaban	Cardiovascular	Yes	=	+		Yes	9,023–9,536	
Apixaban[b]	Cardiovascular	Yes	+	=		Yes	12,300–12,800	
Telaprevir	Infections	Yes		+	=	Yes	10,000–18,000	
Fingolimod	Neurological	No		=	=	Yes	25,000–35,000	
Eribulin	Oncological	Yes	+			No	68,600	Unmet
Boceprevir	Infections	Yes	=	+		Yes	2,909–11,601	
Cabazitaxel	Oncological	Yes	+	=		No	87,500	Unmet
Abiraterone	Oncological	Yes	+	+		Yes	46,800–50,000	Met
Vemurafenib	Oncological	Yes	+	=	=	Yes	44,000–51,800	Met
Pirfenidone	Respiratory	Yes	=	=	=	Yes	24,000	
Ipilimumab	Oncological	Yes	+		=	Yes	42,200	Met
Ipilimumab[b]	Oncological	No	=			Yes	28,600–47,900	Met
Axitinib	Oncological	No	=	=	=	Yes	33,500–52,900	Met
Ruxolitinib	H./oncological	Yes	+	+		No	74,000–149,000	Unmet
Crizotinib	Oncological	Yes[c]	=	=	+	No	50,200–100,000	Met
Pixantrone	H./oncological	No				Yes	22,000	
Aflibercept (Eylea)	Eye	No				Yes	<20,000	
Aflibercept[b] (Eylea)	Eye	No				Yes	12,300–16,800	
Aflibercept (Zaltrap)	Oncological	Yes	+			No	44,000–51,000	Unmet
Dapagliflozin	Metabolic	No				Yes	<20,000	
Ocriplasmin	Eye	Yes	=	+	=	Yes	20,900–30,500	
Bosutinib	Oncological	Yes	+			No	43,000–58,000	Met
Enzalutamide	Oncological	Yes	+	+		Yes	22,600	(Met)
Teriflunomide	Neurological	No	=	=	=	Yes	< 20,000	
Dabrafenib	Oncological	No	=	=	=	Yes	11,000	
Afatinib	Oncological	Yes	+	+	+	Yes	N/A	N/A
Sofosbuvir	Infections	No	=	=		Yes	700–47,600	
Dimethyl fumarate	Neurological	No				Yes	27,700	
Simeprevir	Infections	Yes	=	+		Yes	<20,000	
Mirabegron	Urological	No	=	=	=	Yes	5,270	
Empagliflozin	Metabolic	No				Yes	N/A	
Canagliflozin	Metabolic	No				Yes	N/A	
Nalmefene	Alcohol	No				Yes	5,100	
Sipuleucel-T	Oncological	Yes[c]		=		No	112,000	Unmet
Pomalidomide	H./oncological	Yes	=	+	=	No	50,000–70,000	Unmet

EoL, end of life; GBP, Great British Pound; G-BA, Gemeinsamer Bundesausschuss; H., hematological; HRQoL, health-related quality of life; ICER, incremental cost effectiveness ratio; IQWiG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; N/A, not available; NICE, National Institute for Health and Care Excellence; QALY, quality-adjusted life year; STA, single technology appraisal.

[a]Patient-relevant endpoints/ outcomes: +, significant superiority to comparator; =, superiority to comparator; -, no/ less superiority to comparator.

[b]Additional indication for health technologies.

[c]Health technologies were not recommended with added benefit by IQWiG.

Overall observations, based on the total samples, i.e. including all technology appraisals by NICE and G-BA/IQWiG, were compared descriptively. Because the samples of overall HTAs differ between NICE and G-BA/IQWiG, we identified matching condition-intervention pairs for formal comparative analysis. Matching criteria were that pairs had to address the same intervention for comparable treatment indications and patient population(s). We then compared NICE and G-BA appraisals by capturing the drug ´s main treatment indication (by means of a weighted number of patients for each defined indication of the same intervention), because technologies may differ further with regard to the definitions of patient (sub)groups or selected comparator(s). For detailed information on matched pairs selected for analysis – including information on HTA outcomes, therapeutic area, ICER/ QALY as well as EoL criteria, and patient-relevant outcomes for NICE and G-BA/IQWiG, respectively – see Table 1. To test our results for statistical significance, we compared appraisal outcomes for all matched condition-intervention pairs using the two-tailed chi-square test.

3. Results

During the study period, NICE issued guidance for 88 STAs with 125 subgroups. G-BA completed 105 appraisals with 226 subgroups, whereas IQWiG sliced 105 benefit assessments in 240 subgroups. From these appraisals, we identified 37 matched condition-intervention pairs for NICE and G-BA.

3.1. NICE single technology appraisals

NICE positively evaluated more than 76% (67/88) of all STAs – including orphan drugs. When focusing on patient groups, NICE recommended about 80% (99/125) of all treatment options. ICERs per QALY gained of published STAs were correlated with the recommendation for health technologies ($p < .001$, chi-square test), as an increasing ICER (exceeding £30,000) raises the probability for rejection significantly. NICE recommended all technologies identified with an ICER below £30,000 *(Figure 2)*.

Appraisals for technologies with unspecified ICERs were primarily based on clinical effectiveness as well as uncertainty in the estimated ICER (range). For EoL criteria we found significant correlation between STAs with an ICER exceeding £30,000 and positive recommendations by NICE ($p < .001$, Fisher-Yates test). However, in some cases NICE rejected cancer drugs, although EoL criteria were met (i.e. there was very high uncertainty for the most plausible ICERs of bosutinib and crizotinib; everolimus and pemetrexed were found to be not cost effective compared to best supportive care). The impact of the submitted clinical evidence (in terms of the use of relevant RCTs) was also correlated with technology appraisals, because relevant RCTs were available for all technologies recommended by NICE ($p < .005$, Fisher-Yates test).

3.2. G-BA appraisals and IQWiG dossier assessments

G-BA confirmed additional benefit for more than 60% (64/105) of the assessed health technologies, whereas IQWiG recommended about half (53/105) of those drugs only. By subgroup analysis, G-BA determined added benefit for 90/226 subgroups; IQWiG recommended 75/240 subgroups *(Figure 3)*.

When excluding orphan drugs from statistical analysis, G-BA confirmed added benefit for more than half (45/86) of the drugs (67/203 subgroups), while IQWiG recommended about 40% (34/86) of these technologies (52/217 subgroups). We found significant positive correlation between substantial differences in (at least) one patient-relevant endpoint compared to the ACT and its assessment outcome, for both G-BA appraisals and IQWiG assessments ($p < .001$, chi-square test). With belatacept (urology drug), saxagliptin/metformin (metabolic drug) and sipuleucel-T (cancer drug), we identified only three positively evaluated drug appraisals by G-BA based on surrogate endpoints; in contrast, IQWiG recommended belatacept only. Submitted evidence by manufacturer (in terms of the use of relevant RCTs) is positively correlated with IQWiG dossier assessments as well as with G-BA appraisals ($p < .001$, chi-square test). Compared to G-BA appraisals EBAs by IQWiG are more rigorous, for example, IQWiG rejected all drug assessments that were not based on RCTs.

3.3. Comparative analyses results

3.3.1. Overall observations

Comparing overall HTA results by both agencies, NICE tends to evaluate new drugs more favorably than G-BA *(Figure 4)*. However, differences by therapeutic area apparently exist. Cancer drugs, for example, were more likely to be evaluated positively by G-BA. We also observed differences between G-BA and NICE for the size of defined patient (sub)groups and for the annual costs per patient. Patient (sub)groups defined by G-BA seem to be smaller than patient groups defined by NICE. Costs per patient reported by G-BA and NICE should be considered cautiously, because data may vary with regard to their respective reference base. Our comparative findings suggest that annual treatment costs per patient in Germany tend to be higher than drug acquisition costs in England.

3.3.2. Matched condition-intervention pairs

By comparison, findings for matched pairs were consistent with our overall observations *(Figure 4)*. Of 37 matched pairs, 65% (24/37) differ by HTA outcome. NICE recommended more than 78% (29/37), whereas G-BA confirmed additional benefit for 57% (21/37) ($p < .005$, two-tailed chi-square test). In contrast, only 35% (13/37) of these drug pairs were assessed in an equivalent way; however, no drug was rejected by both agencies.

On the one hand, G-BA confirmed additional benefit for all of the 13/37 drugs (including the orphan respiratory drug pirfenidone), because at least one patient-relevant endpoint had a significant superiority compared to the ACT. Of these 13 drugs,

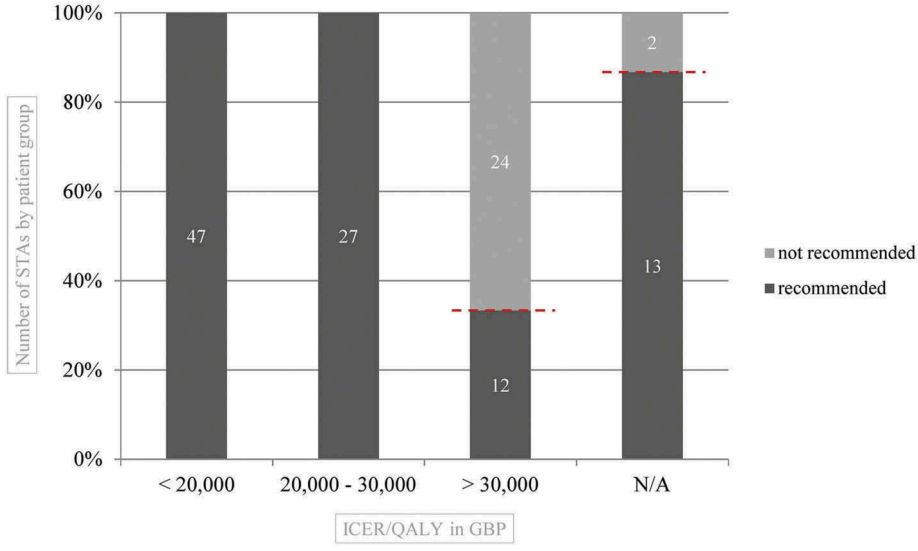

Figure 2. NICE STAs by ICER per QALY gained. GBP, Great British Pound; ICER, incremental cost effectiveness ratio; NICE, National Institute for Health and Care Excellence; QALY, quality-adjusted life year; STA, single technology appraisal.

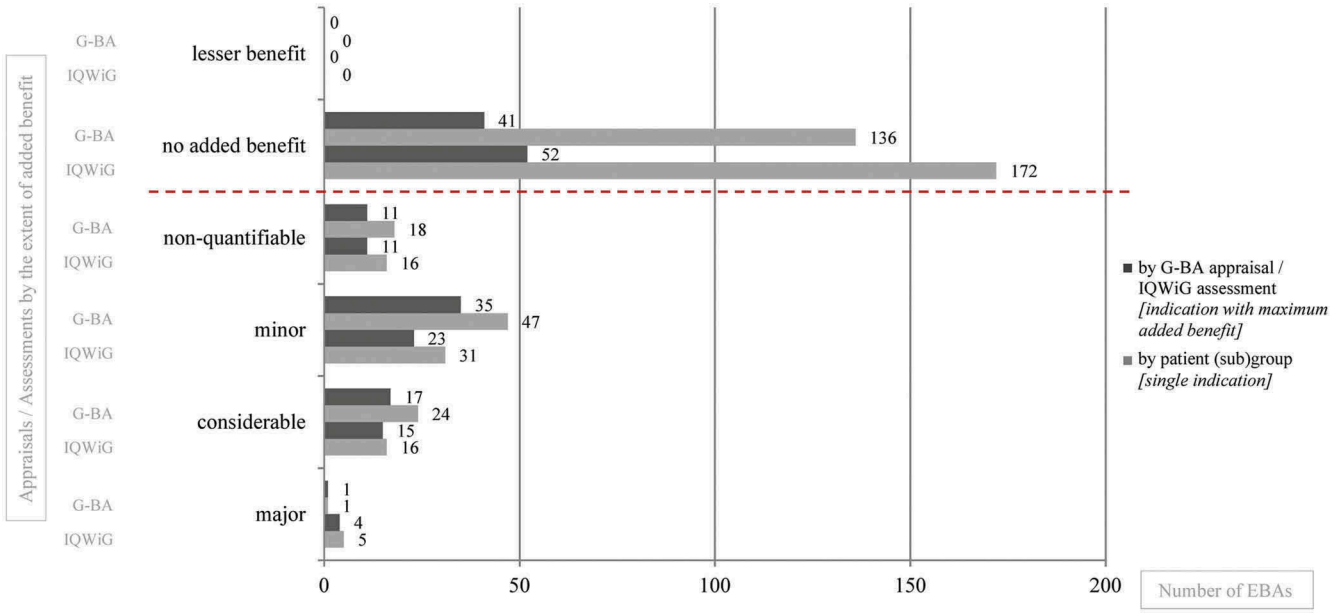

Figure 3. G-BA appraisals and IQWiG dossier assessments by the extent of added benefit. G-BA, Gemeinsamer Bundesausschuss; i.e., id est; IQWiG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen.

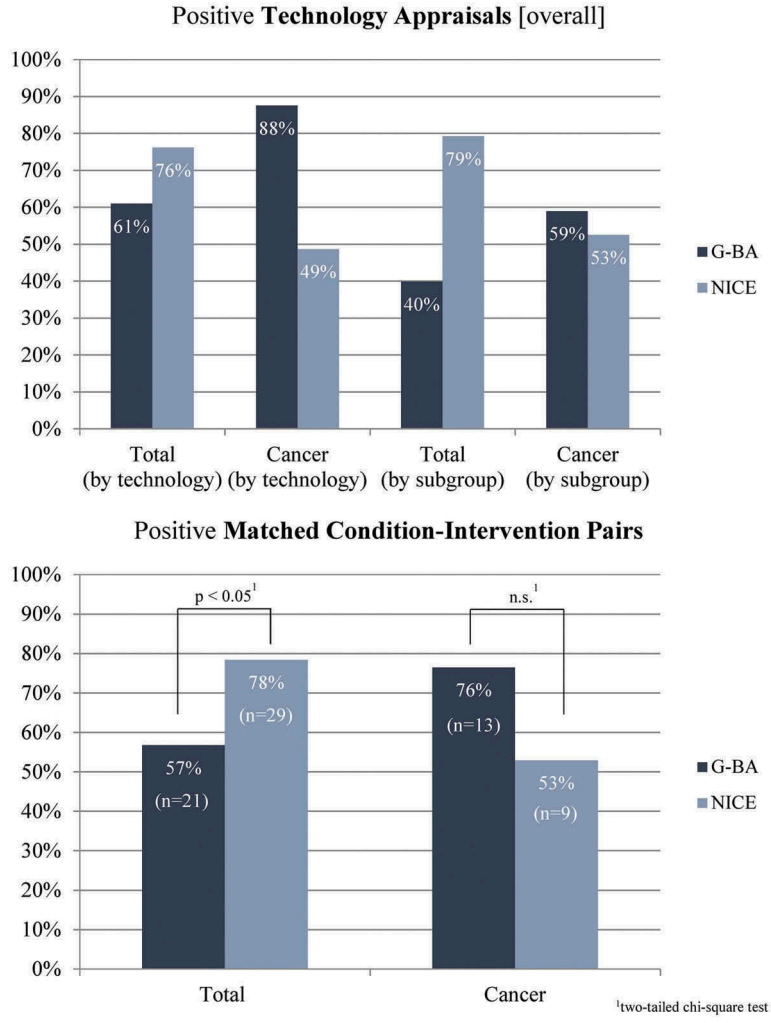

Figure 4. Overview of positive matched condition-intervention pairs and overall technology appraisals (including orphan drugs). G-BA, Gemeinsamer Bundesausschuss; NICE, National Institute for Health and Care Excellence; n.s., not significant.

Table 2. Matched condition-intervention pairs by therapeutic area.

		G-BA decision: Added benefit		NICE guidance: Recommended	
Therapeutic area: Conditions and diseases	Matched pairs	Yes	No	Yes	No
Respiratory	1	1		1	
Eye	3	1	2	3	
Hematological/Oncological	3	2	1	1	2
Cardiovascular	3	3		3	
Infections	4	3	1	4	
Neurological	4		4	4	
Oncological	14	11	3	8	6
Alcohol	1		1	1	
Metabolic	3		3	3	
Urological	1		1	1	
Total	**37**	**21**	**16**	**29**	**8**
Relative share		.57	.43	.78	.22
Total: oncological conditions[a]	**17**	**13**	**4**	**9**	**8**
Relative share		.76	.24	.53	.47

G-BA, Gemeinsamer Bundesausschuss; NICE, National Institute for Health and Care Excellence.
[a]Oncological conditions include hematologic cancer malignancies as well as oncological disorders.

NICE recommended ten drugs due to the estimated ICER, while another three cancer drugs met EoL criteria. On the other hand, benefit assessments of the remaining 24/37 drugs lead to different results; nearly half of the drugs (11/24) were recommended by NICE due to the estimated ICER, whereas G-BA rejected all drugs recommended by NICE due to missing significant superiority in patient-relevant outcomes compared to the ACT. In contrast, G-BA determined additional benefit for 6/24 cancer drugs, including two orphans (bosutinib and ruxolitinib), which were again rejected by NICE (all relevant ICERs exceeded the upper range of £30,000).

Of 29/37 technologies recommended by NICE, nearly 70% (20/29) had an ICER less than £30,000; 3/29 drugs without information on a most plausible ICER were positively evaluated due to clinical effectiveness, and another 5/29 cancer drug recommendations were primarily based on EoL considerations. Retigabine (a neurology drug for the treatment of partial seizures in epilepsy) was recommended due to demonstrated efficacy compared with placebo; however, retigabine was later withdrawn from the market. G-BA confirmed 21/37 matched pairs with added benefit as all of these drugs had (significant) superiority in at least one patient-relevant outcome parameter compared to the ACT.

With 17/37 drugs, interventions for oncological diseases (including hematologic cancer) were identified to be the major therapeutic area (Table 2). Of matched cancer drug appraisals, more than 70% (12/17) differ; while G-BA determined additional benefit for 13/17 (76%) cancer drugs, NICE recommended 9/17 (53%) drugs only ($p = .15$, two-tailed chi-square test).

4. Discussion

Consistent with prior international comparisons of HTA outcomes [21,26–31], we observed substantial differences between NICE in England and G-BA/IQWiG in Germany. The descriptive findings were confirmed by our formal analysis of matched condition-treatment pairs. Despite the small size of our sample of matched pairs, the difference between NICE and G-BA/IQWiG reached statistical significance; NICE issued more positive recommendations for new drugs when compared to confirmation of added benefit by G-BA/IQWiG in Germany. Since both NICE as well as G-BA/IQWiG adhered to their respective official evaluation criteria, our finding may reflect a consequence of the different methodological approaches chosen by both agencies.

Our data reveal differences by therapeutic area. NICE appraisals, for example, were relatively more positive towards treatments for metabolic and neurological disorders. One possible explanation might be that NICE technology appraisals include modelling of long-term outcomes beyond the time horizon documented in clinical studies, whereas G-BA/IQWiG relies more strictly on evidence-based results of RCTs [32,33].

In contrast, interventions for oncological diseases (including hematologic malignancies) were relatively more likely to be evaluated positively by G-BA. Ruof et al. [34] suggest that G-BA in its decisions on cancer drugs primarily focused on survival benefit and disease morbidity (for example, progression-free survival). NICE places relatively more emphasis on cost per patient by applying a cost per QALY threshold, even if the threshold approach may be relaxed under EoL criteria [35]. Accordingly, the probability for a cancer drug to be positively appraised by NICE increased when it met EoL criteria [35].

In addition to cost and clinical effectiveness, some scholars [26–29,31] identified further aspects affecting recommendation decisions, such as process-related and socio-economic factors as well as broader ethical considerations. Similarities in the evaluation process were primarily related to requirements for clinical evidence submitted by pharmaceutical manufacturers. Both NICE and G-BA/IQWiG expect RCTs as the key element of proof of clinical effectiveness. Nonetheless, NICE seems to be relatively more open to accept non-RCTs and indirect comparisons of health technologies [36,37], and thus may offer manufacturers relatively broader options for participation in the assessment and appraisal process [38].

NICE appraisals, however, are largely driven by cost-effectiveness as the dominant assessment criterion [7,10–12,25], even though Dakin et al. [10,11] and Cerri et al. [25] found also that the type of condition as well as the number of RCTs have a significant effect on NICE recommendations. Published retrospective analyses [11,12,25] suggest a cost per QALY threshold well above the officially adopted standard benchmark of £20,000–£30,000/QALY, which correspond to our analysis with ICERs of

positively evaluated technology reaching up to £41,000–£49,000/QALY. Of note, cost per QALY thresholds as a primary decision criterion in health care resource allocation are indeed controversial; while some analyses debate their appropriate range [39,40], others have argued that they are supported neither by theory nor empiric evidence [41,42]. For example, the QALY maximization hypothesis – and hence the very basis for uniform thresholds – was shown to be 'descriptively flawed' [43]. Critical social norms and preferences in the context of health care resource allocation, such as the prevailing concern for the worst off [44,45], are not adequately captured by incremental QALY gains [43,46].

Against this background, German policy-makers opted not to adopt cost per QALY benchmarks; in fact, economic evaluation in official German HTAs has not played a role yet [47]. From the perspective of health economics, this situation is less than ideal, even though solutions overcoming the limitations of the NICE model may be less than straightforward [48]. The German two-stage approach has also been subject to critique on grounds of its perceived rigorousness as well as its weak link to the process of reimbursement price negotiation, which follows the determination of clinical benefit by G-BA [47,49,50]. Interestingly, IQWiG dossier assessments appear to be stricter than the subsequent G-BA decisions [51]. For example, IQWiG concluded that crizotinib and sipuleucel-T (cancer drugs) are not supported by sufficient clinical evidence to confirm added benefit over the respective ACTs, whereas G-BA granted an added clinical benefit rating for both drugs based upon the acceptance of surrogate endpoint benefit.

A main limitation of our study is that the number of matched condition-intervention pairs has been small. Also some patient subgroup definitions varied between NICE and G-BA, so assumptions had been made to compare matched pairs. This means for technologies with patient subgroups that we focused on HTA outcomes for the main patient group to make our results equivalent for the drug´s main treatment indication. While EBAs in Germany were first implemented as part of the AMNOG, we did not consider NICE STAs published before 2011 as well as all MTAs in our analysis. We did also not consider very rare disease treatments for NICE, and we excluded drugs with an orphan status for the statistical analysis of HTA outcomes by G-BA/IQWiG. Furthermore, our analysis focused on primary evaluation criteria only, such as incremental cost-effectiveness or patient-relevant endpoints, which had previously shown to be important determinants of HTA outcomes by NICE [10–12,25] and G-BA [14,16,17], respectively.

The role of other factors, including the budget impact analysis, remains to be established [52]. In a recent study, for example, Mauskopf et al. [53] explored the impact of reimbursement recommendations by NICE on the NHS budget, and found significant correlation between the (potential) budget impact and the degree of reimbursement restrictions. Against this background, further research should explore potential evaluation criteria such as the budgetary impact of new interventions. Focusing on health technologies by therapeutic area might be of relevance to further exploring differences as well as similarities of HTA approaches by NICE and G-BA/IQWiG, and may also support analyses of evaluations of interventions for rare diseases. We intend to continue collecting relevant data on HTAs by both agencies, as a larger sample of matched condition-intervention pairs may provide further insights and statistically significant results.

5. Conclusions

Our study showed that HTA outcomes by the British NICE and the German G-BA differ considerably and in a presumably systematic manner, driven by different primary evaluation criteria. Both agencies followed their official assessment criteria in a consistent manner, apart from well-defined exceptions such as the assessment of orphan drugs in Germany, and EoL considerations, as well as the evaluation program for highly specialised technologies in England. The latter was recently replaced by new decision rules bringing the assessment of ultra-orphan drugs somewhat closer to the standard criteria for cost-effectiveness adopted by NICE.

Our matched pair analysis indicates that overall NICE tends to evaluate new drugs more favourably than G-BA/IQWiG, and thus may be relatively more 'innovation-friendly'. However, our data suggest that new drugs in some therapeutic areas such as cancer were associated with a higher likelihood of a positive appraisal by G-BA/IQWiG. Finally, our results support the hypothesis that different HTA methods contribute to systematic differences in HTA decision-making.

Key issues

- NICE in England tends to evaluate new health technologies more favorably than G-BA/IQWiG in Germany.
- However, new drugs in some therapeutic areas such as cancer or hematology were evaluated more favorably by G-BA/IQWiG.
- Results including all interventions were consistent with the findings reported for matched condition-intervention pairs.
- Our observations confirm that, apart from well-defined exceptions (i.e. orphan drug status in Germany, ultra-orphan diseases and end of life considerations in England), HTAs by both agencies are consistent with their respective official assessment criteria.
- Comparative analysis using the matched-pairs technique support the hypothesis that different HTA methods are associated with systematic differences in HTA decisions.

Acknowledgments

The authors thank Prof. Oliver Schwarz and Dr. Diego Hernández for their support with statistical analysis, as well as Erica Niebauer for editing and proof-reading the manuscript.

Author contribution statement

The research design was developed as a combined effort of all authors. Data extraction as well as descriptive analysis was performed by RS under the supervision of MS. The initial manuscript was drafted by RS and revised by MS. All authors approved the final version to be published; and agree to be accountable for all aspects of the work.

The contribution represents original work that has not been previously published or simultaneously submitted for publication elsewhere.

Data availability statement

The data that support the findings of this study are available from the corresponding author, RS, upon reasonable request.

Funding

This paper was not funded.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Reviewer Disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Banta D. The development of health technology assessment. Health Policy. 2003;63(2):121–132.
2. Lühmann D, Rüther A, Schwarzer R, et al. Grundlagen und Prinzipien von Health Technology Assessment (HTA). In: Perleth M, Busse R, Gerhardus A, et al., editors. Health Technology Assessment. Konzepte, Methoden, Praxis für Wissenschaft und Entscheidungsfindung. 2nd. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft; 2014. p. 1-57.
3. Battista RN, Hodge MJ. The evolving paradigm of health technology assessment: reflections for the millennium. CMAJ. 1999;160 (10):1464–1467.
4. Sorenson C, Chalkidou K. Reflections on the evolution of health technology assessment in Europe. Health Econ Policy Law. 2012;7:25–45.
• **This article reflects the development and evolution of HTA in Europe.**
5. Schlander M. Health technology assessments by the national institute for health and clinical excellence: a qualitative study. New York (NY): Springer; 2007.
6. Taylor R, Mears R. Making decisions at a national level: a NICE experience? In: Kernick D, editor. Getting health economics into practice. Abingdon: Radcliffe Medical Press; 2002. p. 211-222.
7. Schlander M. The use of cost-effectiveness by the National Institute for Health and Clinical Excellence (NICE): no(t yet an) exemplar of a deliberative process. J Med Ethics. 2008;34:534–539.
8. Drummond M, Sorenson C. Nasty or nice? A perspective on the use of health technology assessment in the United Kingdom. Value Health. 2009;12(2):8–13.
9. NICE. Process and methods [PMG9]: guide to the methods of technology appraisal 2013. National Institute for Health and Care Excellence; 2013 [cited 2018 Jul 17]. Available from: https://www.nice.org.uk/process/pmg9
10. Dakin HA, Devlin NJ, Odeyemi IA. "Yes", "No" or "Yes, but"? Multinomial modelling of NICE decision-making. Health Policy. 2006;77:352–367.
11. Dakin H, Devlin N, Feng Y, et al. The influence of cost-effectiveness and other factors on NICE decisions. Health Econ. 2015;24:1256–1271.
•• **Recent study reporting results on cost-effectiveness and other factors influencing NICE decisions.**
12. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Econ. 2004;13:437–452.
• **First analysis to explore NICE´s cost-effectiveness threshold while considering other factors that influence its decisions.**
13. Mason AR, Drummond MF. Public funding of new cancer drugs: is NICE getting nastier? Eur J Cancer. 2009;45(7):1188–1192.
14. IQWiG. Methodenpapier: allgemeine methoden version 5.0. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; 2017 [cited 2018 Jun 28]. Available from: https://www.iqwig.de/download/Allgemeine-Methoden_Version-5-0.pdf
15. Skipka G, Wieseler B, Kaiser T, et al. Methodological approach to determine minor, considerable, and major treatment effects in the early benefit assessment of new drugs. Biom J. 2016;58:43–58.
16. Ivandic V. Requirements for benefit assessment in Germany and England - overview and comparison. Health Econ Rev. 2014;4(12):1–14.
17. BMJV. Sozialgesetzbuch (SGB) Fünftes Buch (V) - Gesetzliche Krankenversicherung - § 35b Kosten-Nutzen-Bewertung von Arzneimitteln. Bundesministerium der Justiz und für Verbraucherschutz; 2007 [cited 2018 Jul 01]. Available from: http://www.gesetze-im-internet.de/sgb_5/__35b.html
18. Höhle-Pasques S, Hankowitz J, Oberender P. Drei Jahre frühe Nutzenbewertung nach §35a SGB V - kritische Würdigung und Lösungsvorschlag. Pharmacoeconomics German Res Art. 2013;11 (1):13–24.
19. Schulz S, Passon A, Kulig M, et al. Studien bei seltenen Erkrankungen: eine deskriptive Analyse abgeschlossener Orphan Drug Bewertungen im Gemeinsamen Bundesausschuss. Gesundheitswesen. 2018;80(12): e54-e61. DOI:10.1055/s-0043-104696.
20. Häussler B, Höer A, de Millas C. Arzneimittel-Atlas 2016. Der Arzneimittelverbrauch in der GKV. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft; 2017.
21. Fischer KE, Heisser T, Stargardt T. Health benefit assessment of pharmaceuticals: an international comparison of decisions from Germany, England, Scotland and Australia. Health Policy. 2016;120 (10):1115–1122.
22. NICE. Guidance and advice list: published. National Institute for Health and Care Excellence; 2015 [cited 2015 Oct 17]. Available from: http://www.nice.org.uk/guidance/published?type=ta
23. G-BA. Informationsarchiv: (Frühe) Nutzenbewertung nach § 35a SGB V. Gemeinsamer Bundesausschuss; 2015 [cited 2015 Oct 18]. Available from: https://www.g-ba.de/informationen/nutzenbewertung/
24. G-BA. Verfahrensordnung des Gemeinsamen Bundesausschusses. Gemeinsamer Bundesausschuss; 2018 [cited 2018 Jul 16]. Available from: https://www.g-ba.de/downloads/62-492-1614/VerfO_2018-03-16_iK-2018-07-05.pdf
25. Cerri KH, Knapp M, Fernandez JL. Decision making by NICE: examining the influences of evidence, process and context. Health Econ Policy Law. 2014;9(2):119–141.
26. Clement FM, Harris A, Li JJ, et al. Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada. JAMA. 2009;302(13):1437–1443.
27. Bae G, Bae EY, Bae SJ. Same drugs, valued differently? Comparing comparators and methods used in reimbursement recommendations in Australia, Canada, and Korea. Health Policy. 2015;119 (5):577–587.
28. Kanavos P, Nicod E, van den Aardweg S, et al. The impact of health technology assessments: an international comparison. Euro Obs. 2010;12(4):1–7.
29. Nicod E, Kanavos P. Commonalities and differences in HTA outcomes: a comparative analysis of five countries and implications for coverage decisions. Health Policy. 2012;108(2–3):167–177.
30. Spinner DS, Birt J, Walter JW, et al. Do different clinical evidence bases lead to discordant health-technology assessment decisions? An in-depth case series across three jurisdictions. Clinicoecon Outcomes Res. 2013;5:69–85.
31. Allen N, Walker SR, Liberti L, et al. Health Technology Assessment (HTA) case studies: factors influencing divergent HTA

reimbursement recommendations in Australia, Canada, England, and Scotland. Value Health. 2017;20:320–328.
32. Akehurst RL, Abadie E, Renaudin N, et al. Variation in health technology assessment and reimbursement processes in Europe. Value Health. 2017;20:67–76.
33. Angelis A, Lange A, Kanavos P. Using health technology assessment to assess the value of new medicines: results of a systematic review and expert consultation across eight European countries. Eur J Health Econ. 2018;19:123–152.
•• **This article reports current practices, processes and policies of value assessment for new medicines in Europe.**
34. Ruof J, Schwartz FW, Schulenburg JM, et al. Early benefit assessment (EBA) in Germany: analysing decisions 18 months after introducing the new AMNOG legislation. Eur J Health Econ. 2014;15(6):577–589.
• **This article analyzes IQWiG dossier assessments as well as G-BA decisions after the introduction of AMNOG in Germany.**
35. Pujolras LM, Cairns J. Why do some countries approve a cancer drug and others don't? J Cancer Policy. 2015;4:21–25.
36. Culyer AJ. NICE's use of cost effectiveness as an exemplar of a deliberative process. Health Econ Policy Law. 2006;1(3):299–318.
37. Lebioda A, Gasche D, Dippel FW, et al. Relevance of indirect comparisons in the German early benefit assessment and in comparison to HTA processes in England, France and Scotland. Health Econ Rev. 2014;4(1):31.
38. Stafinski T, Menon D, Davis C, et al. Role of centralized review processes for making reimbursement decisions on new health technologies in Europe. Clinicoecon Outcomes Res. 2011;3:117–186.
39. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Health Technol Assess. 2015;19(14):1–503.
40. Schlander M, Schwarz O, Hernández D, et al. New estimates of the willingness-to-pay for a statistical life year: a systematic review of the empirical economic literature. Value Health. 2018;21(1):111.
41. Eichler HG, Kong S, Gerth W, et al. Use of cost-effectiveness-analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge? Value Health. 2004;7(5):518–528.
42. Cleemput I, Neyt M, Thiry N, et al. Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. Int J Technol Assess Health Care. 2011;27(1):71–76.
43. Dolan P, Shaw R, Tsuchiya A, et al. QALY maximisation and people´s preferences: a methodological review of the literature. Health Econ. 2005;14(2):197–208.
44. Shah KK, Tsuchiya A, Wailoo AJ. Valuing health at the end of life: A stated preference discrete choice experiment. Soc Sci Med. 2015;124:48–56.
45. Nord E. Beyond QALYs: multi-criteria based estimation of maximum willingness to pay for health technologies. Eur J Health Econ. 2018;19:267–275.
46. Schlander M, Garattini S, Holm S, et al. Incremental cost per quality-adjusted life year gained? The need for alternative methods to evaluate medical interventions for ultra-rare disorders. J Comp Eff Res. 2014;3(4):399–422.
47. Cassel D, Ulrich V. AMNOG auf dem ökonomischen Prüfstand. Funktionsweise, Ergebnisse und Reformbedarf der Preisregulierung für neue Arzneimittel in Deutschland. Gutachten für den Bundesverband der Pharmazeutischen Industrie e.V. (BPI). In: Greiner W, Schreyögg J, Ulrich V, editors. Gesundheitsökonomische Beiträge. 56th ed. Baden-Baden: Nomos; 2015.
48. Caro JJ, Nord E, Siebert U, et al. The efficiency frontier approach to economic evaluation of health-care interventions. Health Econ. 2010;19(10):1117–1127.
49. Fischer K, Stargardt T. Early benefit assessment of pharmaceuticals in Germany: manufacturers´ expectations versus the federal joint committee´s decisions. Med Decis Making. 2014;34(8):1030–1047.
50. Hammerschmidt T. Analyse der AMNOG-Erstattungsbeträge im europäischen Preisumfeld. Gesundh ökon Qual manag. 2017;22 (01):43–53.
51. Bestehorn M, Tunder R. AMNOG - Erste Erfahrungen und mögliche Auswirkungen auf die klinische Forschung. Pharmacoecon German Res Art. 2013;11(1):3–11.
52. Ghabri S, Mauskopf J. The use of budget impact analysis in the economic evaluation of new medicines in Australia, England, France and the United States: relationship to cost-effectiveness analysis and methodological challenges. Eur J Health Econ. 2018;19:173–175.
53. Mauskopf J, Chirila C, Birt J, et al. Drug reimbursement recommendations by the national institute for health and clinical excellence: have they impacted the national health service budget? Health Policy. 2013;110:49–59.