

Accepted Manuscript

Title: Health Benefit Assessment Of Pharmaceuticals: An International Comparison Of Decisions from Germany, England, Scotland And Australia

Author: K.E. Fischer T. Heisser T. Stargardt



PII: S0168-8510(16)30204-4
DOI: <http://dx.doi.org/doi:10.1016/j.healthpol.2016.08.001>
Reference: HEAP 3602

To appear in: *Health Policy*

Received date: 22-3-2016
Revised date: 19-7-2016
Accepted date: 2-8-2016

Please cite this article as: Fischer KE, Heisser T, Stargardt T. Health Benefit Assessment Of Pharmaceuticals: An International Comparison Of Decisions from Germany, England, Scotland And Australia. *Health Policy* <http://dx.doi.org/10.1016/j.healthpol.2016.08.001>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Health Benefit Assessment Of Pharmaceuticals: An International Comparison Of Decisions from Germany, England, Scotland And Australia

FISCHER KE^{1,2}, HEISSER T^{1,3}, STARGARDT T¹

1 Hamburg Center for Health Economics, University of Hamburg, Hamburg, Germany

2 Columbia Business School, Columbia University, New York, USA

3 Ludwig-Maximilians-University Munich, Munich, Germany

Corresponding author: Tom Stargardt, Hamburg Center for Health Economics, Universität Hamburg, Esplanade 36, 20354 Hamburg, Germany, Phone: +494042838-9299, Fax: +494042838-9498, Tom.Stargardt@wiso.uni-hamburg.de

Funding statement: For development of the HCHE AMNOG appraisal database used on reimbursement decisions in Germany, KF and TS have received funding by the German Research Foundation (grant numbers FI 1950/2-1 and STA 1311/2-1). KF is funded with a research grant from the Federal Ministry of Education and Research in Germany (grant number BMBF 01EH1101A). TS was involved in receiving the grant.

Conflict of interest: None to declare.

Acknowledgements: None.

Keywords: Decision Making; Priority Setting; Early Benefit Assessment; Reimbursement of pharmaceuticals

Word count: 3,657

Highlights

- We identified substantial disagreement between the FJC and NICE, SMC and PBAC.
- FJC and each agency agreed in 40%, 47.6% and 48.7% of ratings.
- Agreement improved moderately when comparing decisions on effectiveness only.
- Generally, the FJC tends to appraise stricter than NICE.

Abstract

Background: Little is known on the performance of the newly introduced health benefit assessment process, AMNOG, in Germany compared to other health technology assessment agencies.

Objective: We analysed whether decisions of the German Federal Joint Committee (FJC) deviate from decisions of the UK National Institute for Health and Care Excellence (NICE), the Scottish Medicines Consortium (SMC) and the Australian Pharmaceutical Benefits Advisory Committee (PBAC).

Methods: We analysed decisions made for comparable patient subgroups by the four agencies between 2011 and 2014. First, decisions were compared (a) by their final outcome, i.e. whether a health benefit was identified, and (b) by the agencies' judgement on comparative effectiveness. Subsequently, we partially explored reasons for differences between HTA agencies.

Results: From the 192 FJC decisions, we identified 55 that overlapped with NICE, 166 with SMC and 119 with PBAC. FJC agreed with NICE in 40% in final outcome (Cohen's kappa=-0.13). Similar results were obtained for FJC and SMC (47.6%, kappa=0.03) and FJC and PBAC (48.7%, kappa=0.07). Agreement increased when comparing judgements based on comparative effectiveness only. However, the FJC's final decision was positive only in 43.6%, 39.2% and 44.5% of the patient subgroups, as opposed to 74.5% (NICE), 68.7% (SMC), and 68.9% (PBAC), respectively.

Conclusion: We show that the FJC - an agency relatively new in structurally assessing the health benefit of pharmaceuticals - deviates considerably in decisions compared to other HTA agencies. Our study also reveals that the FJC tends to appraise stricter than NICE.

1. Introduction

Fourth-hurdle decision making helps to decide on a new pharmaceutical's coverage and reimbursement within a health care system. It is called 'fourth-hurdle' because the pharmaceutical has already passed three hurdles to achieve market authorization thereby demonstrating its safety, efficacy and quality [1,2]. Given the need to allocate scarce resources and to contain pharmaceutical expenditure, many countries have established fourth-hurdle decision making to assess and appraise technologies over the last two decades [3]. Among them are Australia, Belgium, Canada, England and Wales, Scotland, Sweden and the Netherlands [4]. The general aim is to assess the trade-off between health benefit and a pharmaceutical's cost. Despite being Europe's largest pharmaceutical market in terms of sales volume, Germany was a late-mover to implement the fourth hurdle in January 2011, by the Pharmaceutical Market Restructuring Act (AMNOG).

Within three months after market launch, all newly introduced pharmaceuticals are evaluated based on their added benefit over a comparator, the so-called Early Benefit Assessment (EBA) [5]. By law, manufacturers are obliged to submit a dossier to the Federal Joint Committee (FJC). Within six months after submission, the FJC performs the appraisal. Whilst the final decision is with the FJC, the Institute for Quality and Efficiency in Health Care (IQWiG) is, by convention, commissioned for a preliminary assessment. While IQWiG assesses the evidence submitted by the manufacturer in the first place, FJC is responsible for the final decision after a separate assessment of the evidence. It has been shown that FJC tends to soften IQWiG's decisions [6]. If an added benefit is approved by FJC, in a separate stage, the manufacturer and the Federal Association of Sickness Funds negotiate a price within another six months. Pharmaceuticals that do not show an added benefit become subject to reference pricing or other reimbursement restrictions.

To date, evidence on the German system focusses on discussions on the benefits and limitations of the AMNOG reform itself, the outcomes of the first wave of EBAs and the agreement between manufacturers, IQWiG and FJC [6–13]. However, little is known about how the FJC's judgements compare to other health technology assessment (HTA) agencies. First international comparisons have revealed differences in the process and provided qualitative overviews of decisions by therapeutic areas and prices [14,15]. Other studies examine consideration of indirect comparisons [10] or quality of life [16] in appraisals. This is of relevance to both the pharmaceutical industry and health policy makers.

While pharmaceutical product development aims to cover multiple health care markets, regulation ideally follows country specific preferences. This is why there are varying preferences towards the process and methods of evaluating new pharmaceuticals [17]. For this reason, the institutions that have emerged share some common features, but also differ in others. For the final decision, important criteria in many systems are the appraisal of comparative effectiveness, i.e. the appraisal of ‘*clinical information on the relative merits or outcomes of one intervention in comparison to one or more others*’ [18] and, cost-effectiveness that analyses a substance’s benefits in face of its cost. While cost-effectiveness is not considered in the FJC process, it is a common criterion of nearly all HTA agencies in the appraisal process. Comparative research thus helps (a) pointing out areas of disagreement between agencies when performing the same or similar tasks, (b) identifying and explaining drivers for deviation in outcomes and (c) improving decision-making processes.

Previous research has analysed the fourth hurdle through various means [19,20]. First quantitative approaches have focussed on the final appraisal, i.e., the resulting decision that may rest on varying criteria across HTA agencies and the appraisal of comparative effectiveness [21,22]. Qualitative approaches have explored possible reasons for variations in decisions by variation in the decision-making criteria and the reasons for differences in HTA including the varying interpretation in underlying uncertainty of the evidence [23]. Such approaches typically cover a smaller sample of decisions and specific product categories that allows in-depth analysis of the interpretation of available evidence by all types of sub-criteria and including the full complexity of decision-making. In this study, we focus on the final decision and the assessment of comparative effectiveness as these outcomes determine the degree of implementation in the health system after the decision and constitute what is perceived by stakeholders first.

The objective of this study was thus to compare the decisions of the German FJC with three other HTA agencies. We chose the English National Institute for Health and Care Excellence (NICE), the Scottish Medicines Consortium (SMC), and the Australian Pharmaceutical Benefits Advisory Committee (PBAC) as comparator HTA agencies. As a first step, we analysed the decisions made jointly by the FJC and the other agencies between 2011 and 2014. Finally, we partially explored drivers for deviances in outcome.

2. Comparator Agencies

Evaluations of benefit in all four agencies are conducted in two stages, as separate institutions for assessment and appraisal are involved. FJC, NICE and PBAC appraisals are comprehensive, while SMC conducts a ‘rapid early review’ [24]. The trigger for the benefit assessment process differs: FJC and SMC appraise all newly licensed pharmaceuticals, PBAC requires manufacturers to actively seek reimburse-

ment. NICE reviews all cancer drugs and most new indications or new entities as it sees fit. However, there is no formal requirement to review new drugs / indications to receive market access. This also implies differences in the timing of the assessment. FJC, PBAC and SMC appraise new entities early in a drug's post-development life cycle, while the time frame for NICE appraisals may vary. Furthermore, the consequences for pharmaceuticals' pricing and reimbursement differ. A negative decision by NICE, SMC and PBAC will exclude a drug from reimbursement. This may be based on unfavourable comparative effectiveness or, if health benefits are present, cost-effectiveness. A negative decision of the FJC that solely rests on considerations of comparative effectiveness will 'only' impact reimbursement prices. Thus, the results from the appraisal of comparative effectiveness have differing consequences.

With respect to the type of evidence taken into account in the decision-making process, all agencies use clinical evidence for their appraisals. While the FJC's assessment is totally limited to clinical evidence and only evaluates comparative effectiveness [6], NICE, SMC and PBAC follow a 'value-for-money' approach in addition and their assessment includes cost-effectiveness [25–27]. Given that the final decision may vary across HTA agencies because of the diverging criteria between FJC (i.e. comparative effectiveness) and NICE, SMC, and PBAC (i.e. cost-effectiveness, cost minimization), besides the final decision, we compare the appraisals of comparative effectiveness in this study. Technologies which are not considered effective in terms of health benefit are typically dominated in terms of cost-effectiveness. Besides, cost-effectiveness considerations are subject to strategic pricing behaviour of manufacturers. Table 1 provides an overview of the key differences and similarities of the four HTA agencies that were included in our analysis. Details have been described elsewhere [24,28–30].

- Table 1 about here -

3. Methods

3.1 Study Setting

To begin with, we screened all EBAs of the FJC concluded between January 2011 and December 2014 and extracted data on decisions on the patient-subgroup level. In case of revisions, we only considered the latest decision. Decisions that were ongoing, had no status, were discontinued or involved pharmaceuticals that were suspended from AMNOG were excluded. Subsequently, we extracted all data on decisions of NICE, SMC and PBAC appraisals for which a corresponding FJC-decision on patient subgroup-level existed.

For FJC decisions and information on the corresponding EBAs, we used a database developed by the Hamburg Center for Health Economics (HCHE). The database builds on publicly available data from the FJC's website (www.g-ba.de) and has been used in previous research [6]. For NICE, SMC and PBAC decisions and underlying benefit assessments, data were extracted from public summary documents available from the agencies' websites (NICE: www.nice.org.uk ; SMC: www.scottishmedicines.org.uk ; PBAC: www.pbs.gov.au).

Data extraction was performed in several consecutive steps. In a first step, we captured the overlap of decisions between FJC and the other HTA agencies. To ensure that decisions and extracted data are comparable in their scope, we defined a set of variables that capture the substance, main indication, patient subgroup and whether an appraisal was performed at subgroup level by the respective agency. This process was pre-tested with a small set of decisions and then performed consistently across the three other HTA agencies. In a second step we matched decisions of the other HTA agencies if they covered the patient subgroup considered or a wider group of patients as the patient subgroups of FJC are often more refined. Finally, we extracted the final decision and comparative effectiveness considerations through two separate reviewers. All conflicts were resolved by discussion between the three authors.

To ensure that the scope of the decisions was similar across HTA agencies, we captured the pharmaceutical's main indication. As HTA agencies frequently specify a number of patient subgroups by clinical criteria, we also ensured that these were equal between the FJC and the other three agencies.

3.2 Outcome Measurement

The final decision of the four HTA agencies was our outcome variable to analyse agreement. To ensure comparability with the other HTA agencies, we used binary coding ('added benefit', 'no added benefit') for the FJC decisions. The FJC categories 'major', 'considerable', 'minor', or 'non-quantifiable added benefit' were considered for the category 'added benefit'. The other two categories, 'no evidence of added benefit' and 'less benefit', were combined to 'no added benefit'.

As the FJC decides on comparative effectiveness only, we collected data on NICE, PBAC and SMC judgement's in two ways: (a) We collected data on the final decision of the other HTA agencies, i.e., cost-effectiveness. Again, we differentiated between positive ratings (i.e., recommended / recommended with restriction for NICE and PBAC and accepted / accepted with restriction for SMC) and negative ratings (i.e., not recommended for NICE, SMC and rejected for PBAC). (b) On a separate basis, we collected data on the HTA agencies' rating on a pharmaceutical's comparative effectiveness only. We differentiated between 'better' if the HTA agency stated that the pharmaceutical is superior in comparative effectiveness or 'similar or not proven'. We consider this categorization to be equivalent to the FJC's final decision

(added benefit / no added benefit). Due to the extraction of judgements being prone to subjectivity, the rating of comparative effectiveness was extracted by two independent reviewers. Any disagreement was resolved by discussion. Interrater reliability was substantial, with a Cohen's kappa of 0.78 for NICE, 0.85 for SMC and 0.82 for PBAC ratings of comparative effectiveness.

To conclude in a rating of health benefit, the FJC appraises health outcomes in four categories of clinical endpoints – mortality, morbidity (of a disease's medical conditions), adverse events and quality of life – in every patient subgroup. We captured the other HTA agencies' ratings of clinical endpoints as for the FJC. Using this information, we created dummy variables to express agreement between the FJC and the other HTA agencies in endpoints. We defined agreement as positive (FJC: added benefit; NICE / SMC / PBAC: better) or negative (FJC: no added benefit; NICE / SMC / PBAC: similar or not proven) rating in each of the four endpoint categories. As these variables were again prone to subjectivity, data were extracted independently by two reviewers. Disagreement was resolved by discussion. Across endpoints, interrater reliability varied. Highest agreement was for mortality (mean kappa: 0.79) and lowest was for morbidity (mean kappa: 0.30).

In addition, we captured (a) the difference in the number of RCTs available to the FJC and each of the other HTA agencies, (b) the number of months between the FJC's decision and the decision of each of the other HTA agencies and (c) whether the comparator(s) of each of the three HTA agencies include(s) (potentially among others) the comparator used by the FJC.

3.3 Statistical Analysis

We performed descriptive analyses by contrasting the FJC decisions on the level of patient subgroups in two ways. We compared (a) the corresponding final decision of FJC and the other HTA agencies (i.e. e.g. added benefit / no added benefit vs. recommended / not recommended) and (b) only the corresponding judgements on comparative effectiveness (i.e. added benefit / no added benefit vs better / similar or not proven). For each comparison, we analysed the agreement between FJC and each of the other HTA agencies. Agreement was quantified by calculating Cohen's kappa, a measure to determine whether agreement between two raters is by chance [31]. Kappa ranges from -1 to 1; with a value less than or close to 0 indicating that agreement is due to chance. A value of 1 indicates perfect agreement whilst the -1 reflects perfect disagreement. All analyses were performed using R version 2.15.3.

4. Results

4.1 Description of Study Sample

118 EBAs had been concluded by FJC within the timeframe of our study. Of these, we excluded 19 according to our exclusion criteria. For the remaining 99 EBAs that included in total 192 patient subgroups, we identified corresponding appraisals for NICE (35), SMC (66) and PBAC (66), respectively. These represented 55 (NICE), 166 (SMC) and 119 (PBAC) patient subgroups (see Figure 1). For the comparison FJC-NICE, the majority of patient subgroups were in the disease categories ‘neoplasms’ (44%), ‘diseases of the nervous system’ (15%) and ‘diseases of the circulatory system’ (13%). For FJC-SMC and FJC-PBAC, top three disease categories were ‘infectious and parasitic diseases’ (FJC-SMC: 20%, FJC-PBAC: 21%), ‘neoplasms’ (FJC-SMC: 24%, FJC-PBAC: 26%) and ‘endocrine nutritional and metabolic diseases’ (FJC-SMC: 28%, FJC-PBAC: 22%). Furthermore, there were a number of patient subgroups in each comparison where orphan drugs have been indicated for: 4 patient subgroups in the comparison between NICE and FJC, 20 in the FJC-SMC and 13 in the FJC-PBAC comparison, respectively.

- Figure 1 about here -

4.2 Agreement in Decisions

Agreement in final decision (added benefit / no added benefit vs. recommended / not recommended) between the FJC and the three other HTA agencies was low (Table 2): FJC and NICE agreed in only 40% (22 of 55 patient groups) of their ratings. Results were similar for the comparisons FJC-SMC and FJC-PBAC, with 47.6% (79 of 166) and 48.7% (58 of 119) agreement, respectively. Coefficients of Cohen’s Kappa also indicated low agreement with -0.13, 0.03 and 0.07 for the comparison of FJC with NICE, SMC and PBAC, respectively.

Agreement improved moderately when comparing FJC decisions with the other agencies’ judgements on comparative effectiveness only: FJC agreed with NICE in 52.7%, with SMC in 64.5%, and with PBAC in 69.7% of patient subgroups. Again, estimates of Cohen’s Kappa indicate low agreement, with 0.11, 0.26 and 0.38, respectively. Across all comparisons of agreements, the results reveal a consistent pattern.

The FJC rated substantially fewer patient subgroups positive in its final decision than the other three HTA agencies. The FJC decided positive only in 43.6%, 39.2% and 44.5% of the patient subgroups, as opposed to 74.5% (NICE), 68.7% (SMC), and 68.9% (PBAC), respectively.

- Table 2 about here -

4.3 Agreement in Endpoints and Choice of Comparator

To compare agreement in endpoints, we excluded those appraisal decisions where manufacturers did not submit a dossier (FJC: 19 patient subgroups; NICE: 3; SMC: 26; PBAC: 2) or data on some of our variables was not reported in public summary documents (12 patient subgroups for FJC-SMC, 4 patient subgroups for FJC-PBAC). Thus, our results were based on 48 (FJC-NICE), 122 (FJC-SMC) and 109 (FJC-PBAC) patient subgroups.

Agreement in endpoints between the agencies was highest for adverse events (92% for FJC-NICE, 77% for FJC-SMC and 77% for FJC-PBAC) and quality of life (85% for FJC-NICE, 89% for FJC-SMC and 96% for FJC-PBAC) followed by mortality (71% for FJC-NICE, 83% for FJC-SMC and 83% for FJC-PBAC). However, one has to keep in mind that these were also the categories where ‘no additional benefit’ or ‘similar / not proven’ was scored more often, thus leading to agreement in a negative decision. For morbidity, FJC and the other agencies agreed least often (52% for FJC-NICE, 60% for FJC-SMC and 64% for FJC-PBAC).

The comparator(s) used by the FJC was also allowed by the three other HTA agencies (potentially among others) in 71% of decisions for NICE, 56% for SMC and 50% for PBAC, respectively.

5. Discussion

This study contributes to the research on coverage and reimbursement decision-making by examining whether and to what extent decisions of the FJC deviated from those of other HTA agencies. Our findings further support developing hypotheses on what may drive FJC decisions as opposed to other HTA agencies. This also allows initiating a discussion whether the outcomes of FJC’s appraisal process are in accordance with the preferences of the German population. In particular, as decisions are put in a broader, international context.

We identified substantial disagreement in final decisions between the FJC and the other HTA agencies. This finding is not completely unexpected, given the differences in the HTA agencies’ mandates, characteristics and their decision-making processes. Reviewing the existing literature to similar international comparisons, several authors compared methodological aspects of health benefit assessment procedures [4,32–34]. Generally, our findings are in line with other studies, which also showed that HTA agencies disagree in their decisions. For example, Nicod and Kanavos report poor to moderate agreement in comparing NICE, PBAC, SMC, the Canadian Common Drug Review and the Swedish Dental and Pharmaceutical Benefits Board [37]. Still, it is surprising that the disagreement persists when the comparison between the HTA agencies is limited to the judgement of comparative effectiveness only.

As orphan drugs are frequently considered in special in decision processes [39,40], we accounted for potential differences by excluding decisions on orphan drugs across HTA agencies as a sensitivity analysis. Results remained robust. The coefficients of Cohen's Kappa for the agreement in final decision (FJC-NICE: -0.06; FJC-SMC: 0.07; FJC-PBAC: 0.006) and comparative effectiveness (FJC-NICE: 0.11; FJC-SMC: 0.26; FJC-PBAC: 0.40) were similar to the full sample.

The extent of disagreement may, however, also raise concerns given the overlap in the evidence that was submitted to all agencies. The absolute difference in the number of RCTs between FJC and the other HTA agencies was, on average, one study (1.02 studies for FJC vs. NICE, 1.23 for FJC vs. SMC and 1.05 for FJC vs. PBAC). However, when contrasting the final decisions between NICE, SMC and PBAC based on our sample, a high level of agreement was only found between NICE and SMC (70% agreement based on 50 subgroups, Cohen's Kappa: 0.63). The level of agreement between NICE and PBAC (61% based on 41 patient subgroups, Cohen's Kappa: 0.23) and between SMC and PBAC (50% based on 111 patient subgroups, Cohen's Kappa: 0.11) was similar to the comparisons to the FJC. This suggests that agreement between HTA agencies is rather the exception than the rule.

Part of the differences in decision-making may be related to the impact of an agencies' decision on patient access. Whereas a negative decision by NICE, SMC as well as PBAC excludes a drug from reimbursement and thus automatically restricts patient access, the consequences of FJC decisions are not that severe. In fact, when the FJC does not find an 'added benefit', the drug will still be reimbursed although at a (potentially) much lower price. This means that the drug will be subject to other regulations such as reference pricing. However, unless the manufacturer withdraws its pharmaceutical from the market, FJC's decisions do not compromise patient access (just manufacturer's profits). This could explain the more strict line in decision-making of the FJC compared to the three other HTA agencies.

We also identified explanations that are rooted in deviations of the methodological framework of the agencies: (1) Agencies differ in accepting endpoints such as recognising the surrogate endpoint progression-free-survival. Another example is to prefer disease-specific mortality over overall mortality as endpoint or vice versa. (2) Agencies differ in the choice of comparator(s). For example, when deciding on the added benefit for Apixaban for the treatment of atrial fibrillation, the FJC viewed Vitamine-K-Antagonists as comparators whereas NICE and SMC defined Rivaroxaban and Dabigatran in addition to Vitamine-K-Antagonists as comparators. PBAC even went further by adding acetylsalicylic acid to the list of comparators. This will render the same evidence useful for the appraisal of one agencies whereas it cannot be used in the appraisal of another agency. (3) There are differences in handling lack of evidence. An example may arise from the decisions made on Apixaban for the treatment of pulmonary embolism and for the prevention of recurrent venous thromboembolism. Whereas NICE and FJC both agree that evidence was

only presented for short term prophylaxis, NICE concluded that there was no evidence of a difference in the effectiveness to the two other agents in that class (rivaroxaban and dabigatran) [41]. For rivaroxaban it had judged earlier on the absence of evidence: *'the committee accepted that there was no [...] reason why the effects [...] would not be maintained over the longer term'* [42], whereas FJC concluded that the manufacturer had not provided data and thus no added benefit was found. (4) Strategic behaviour may strain outcomes of initial submissions. As manufacturers submit a price suggestion in Australia upon submission, PBAC is likely to have taken a tougher line on aspects of evidence in part to reduce prices on resubmission. (5) The systems included are at different maturity stages in the implementation of the fourth hurdle. Having started appraisal of health benefit not until 2011, FJC has about 10 years less experience in evaluating pharmaceuticals on a regular basis compared to NICE.

Other authors named agency processes, attitudes to the added benefit of a new pharmaceutical within an established therapeutic class including clinical and economic evidence, differences in the extent of required information, date of decision, the rigour of the appraisal, characteristics of the reimbursement system and the country's socioeconomic situation as reasons for differences in decision-making [29,35,36].

Our study has several methodological limitations. Most importantly, our analyses only cover components and decisions that are present across at least two decision-making processes. For the time frame of our study, we included about 32% / 39% / 20% of NICE's / SMC's / PBAC's total number of technology appraisals. Restricting the decisions to those where a paired comparison is possible may not show the other HTA agencies' typical decision-making behaviour. However, compared to similar studies published in the literature, our results appear to be valid. For example, for the patient subgroups included in our study, the proportion of positive final decisions of FJC, NICE, SMC, and PBAC was 39-45%, 75%, 69% and 69%, respectively. Those found by other authors ranged from 39-50% for FJC [6,13], 71-87% for NICE [29,35,43-45], 57-68% for SMC [3,29,46], and 62-74% for PBAC [29,35]. Of course, our results could also be influenced by the process of data extraction.

Although we carefully chose to compare patient subgroups between FJC and the other HTA agencies, these do not always represent perfect matches. In some appraisals, NICE, SMC or PBAC appraised the substance in the same indication, but not all patient subgroups were defined exactly the same as by the FJC. For example, for Sofosbuvir in the treatment of chronic hepatitis C infection, the FJC defined patient subgroups according to treatment experience and genotype. One patient subgroup specifically corresponded to patients with HIV-coinfection. The PBAC, however, likewise referred to treatment experience and genotype, but did not comment on HIV-coinfections separately. Moreover, the exact endpoints considered and comparator against which a substance is appraised may differ across HTA agencies. Thus, decisions may differ not by diverging views on comparative effectiveness, but based on varying reference points,

i.e. definitions of comparators and endpoints. However, the selection of comparators typically is not exogenous in the appraisal process.

Another aspect that influences the degree of implementation of a product is the extent of uncertainty underlying the decision. For this purpose, FJC provides an explicit categorical rating to display the degree of uncertainty of the evidence considered. A very detailed analysis of reasons for deviances in outcomes has been conducted by Nicod and Kanavos based on two assessments in England, Scotland, Sweden and France [23]. The study shows that NICE rejected one of the substances because of uncertainty in cost-effectiveness.

6. Conclusions

In modern health care systems, fourth hurdle decision-making plays a key role in assuring sustainable use of resources to guarantee that patients will receive adequate and likewise effective treatments in the long run. The present study showed that the German FJC, that is relatively new to the field of health benefit assessment of pharmaceuticals, considerably deviates in its outcomes from established HTA agencies, even when only considering aspects of comparative effectiveness. While this can be attributed to differences in the agencies' mandates, characteristics as well as the consequences of a negative decision for patient access, our study reveals that the FJC tends to appraise stricter in terms of comparative effectiveness than NICE.

References

1. Taylor RS, Drummond MF, Salkeld G, Sullivan SD. Inclusion of cost effectiveness in licensing requirements of new drugs: the fourth hurdle. *BMJ*. 2004 Oktober;329(7472):972–5.
2. Rawlins MD. Crossing the fourth hurdle: Crossing the fourth hurdle. *Br. J. Clin. Pharmacol.* 2012 Jun;73(6):855–60.
3. Charokopou M, Majer IM, Raad J de, Broekhuizen S, Postma M, Heeg B. Which Factors Enhance Positive Drug Reimbursement Recommendation in Scotland? A Retrospective Analysis 2006–2013. *Value Health*. 18(2):284–91.
4. Levy AR, Mitton C, Johnston KM, Harrigan B, Briggs AH. International comparison of comparative effectiveness research in five jurisdictions: insights for the US. *PharmacoEconomics*. 2010;28(10):813–30.
5. Federal Joint Committee, editor. Chapter 5: Benefit Assessment of pharmaceuticals according to s. 35a SGB V [Internet]. [cited 2015 Jun 29]. Available from: [http:// www.english.g-ba.de/downloads/17-98-3042/ Chapter5-Rules-of-Procedure-G-BA.pdf](http://www.english.g-ba.de/downloads/17-98-3042/Chapter5-Rules-of-Procedure-G-BA.pdf)
6. Fischer KE, Stargardt T. Early Benefit Assessment of Pharmaceuticals in Germany: Manufacturers' Expectations versus the Federal Joint Committee's Decisions. *Med. Decis. Making*. 2014 Nov 1;34(8):1030–47.
7. Schlette S, Hess R. Early benefit assessment for pharmaceuticals in Germany: lessons for policy-makers. *Issue Brief Commonw. Fund.* 2013 Oct;29.
8. Gerber A, Stock S, Dintsios C-M. Reflections on the Changing Face of German Pharmaceutical Policy. *PharmacoEconomics*. 2011 Jul 1;29(7):549–53.
9. Henschke C, Sundmacher L, Busse R. Structural changes in the German pharmaceutical market: Price setting mechanisms based on the early benefit evaluation. *Health Policy*. 2013 Mar;109(3):263–9.
10. Lebioda A, Gasche D, Dippel F-W, Theobald K, Plantör S. 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 Dec;4(1):31.
11. Ivandic V. Requirements for benefit assessment in Germany and England - overview and comparison. *Health Econ. Rev.* 2014 Dec;4(1):12.
12. Hörn H, Nink K, McGauran N, Wieseler B. Early benefit assessment of new drugs in Germany - results from 2011 to 2012. *Health Policy Amst. Neth.* 2014 Jun;116(2–3):147–53.
13. Ruof J, Schwartz FW, Schulenburg J-M, Dintsios C-M. Early benefit assessment (EBA) in Germany: analysing decisions 18 months after introducing the new AMNOG legislation. *Eur. J. Health Econ. HEPAC Health Econ. Prev. Care*. 2014 Jul;15(6):577–89.

14. Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie. Frühe Nutzenbewertung neuer Arzneimittel in Deutschland: 2011-2014. Freund M, Lüftner D, Wilhelm M, editors. 2015.
15. Szerb A, Kanavos P. Health Technology Assessment of Cancer Drugs in France and Germany: Commonalities and Differences in the Value Assessment of Medical Technologies. LSE Health Work. Pap. 2015;(43).
16. Lohrberg D, Augustin M, Blome C. The definition and role of quality of life in Germany's early assessment of drug benefit: a qualitative approach. *Qual. Life Res.* 2015 Aug 5;1–9.
17. Drummond M. Twenty years of using economic evaluations for drug reimbursement decisions: what has been achieved? *J. Health Polit. Policy Law.* 2013 Dec;38(6):1081–102.
18. Chalkidou K, Tunis S, Lopert R, Rochaix L, Sawicki PT, Nasser M, et al. Comparative Effectiveness Research and Evidence-Based Health Policy: Experience from Four Countries. *Milbank Q.* 2009 Jun 1;87(2):339–67.
19. Fischer KE, Leidl R. Analysing coverage decision-making: opening Pandora's box? *Eur. J. Health Econ.* 2014;15(9):889–906.
20. Fischer KE. A systematic review of coverage decision-making on health technologies—Evidence from the real world. *Health Policy.* 2012;107(2–3):218–30.
21. Bending M, Hutton J, McGrath C. A comparison of pharmaceutical reimbursement agencies' processes and methods in France and Scotland. *Int J Technol Assess Health Care.* 2012;28:187–94.
22. Dakin H, Devlin N, Feng Y, Rice N, O'Neill P, Parkin D. The Influence of Cost-Effectiveness and Other Factors on Nice Decisions. *Health Econ.* 2014 Sep 1;n/a-n/a.
23. Nicod E, Kanavos P. Developing an evidence-based methodological framework to systematically compare HTA coverage decisions: A mixed methods study. *Health Policy.* 2016 Jan;120(1):35–45.
24. Cairns J. Providing guidance to the NHS: The Scottish Medicines Consortium and the National Institute for Clinical Excellence compared. *Health Policy.* 2006 Apr;76(2):134–43.
25. NICE. Guide to the processes of technology appraisal [Internet]. 2014. Available from: <http://www.nice.org.uk/article/pmg19>
26. SMC. Working with SMC – A Guide for Manufacturers [Internet]. 2014 [cited 2014 Jun 7]. Available from: https://www.scottishmedicines.org.uk/files/submissionprocess/Working_with_SMC_July_2014.pdf
27. PBAC. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee [Internet]. 2013 [cited 2015 Jun 7]. Available from: <http://www.pbac.pbs.gov.au/content/information/printable-files/pbacg-book.pdf>
28. Ford JA, Waugh N, Sharma P, Sculpher M, Walker A. NICE guidance: a comparative study of the introduction of the single technology appraisal process and comparison with guidance from Scottish Medicines Consortium. *BMJ Open.* 2012;2(1):e000671.
29. Kanavos P, Nicod E, van den Aardweg S, Pomedli S. The impact of health technology assessments: an international comparison. *Euro Obs.* 2010;12(4):1–7.

30. Lopert R, Elshaug AG. Australia's "fourth hurdle" drug review comparing costs and benefits holds lessons for the United States. *Health Aff. Proj. Hope*. 2013 Apr;32(4):778–87.
31. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* 1960 Apr 1;20(1):37–46.
32. Stafinski T, Menon D, Philippon DJ, McCabe C. Health Technology Funding Decision- Making Processes Around the World. *PharmacoEconomics*. 2011 Jun;29(6):475–95.
33. Kleijnen S, George E, Goulden S, d'Andon A, Vitre P, Osińska B, et al. Relative effectiveness assessment of pharmaceuticals: similarities and differences in 29 jurisdictions. *Value Health J. Int. Soc. Pharmacoeconomics Outcomes Res.* 2012 Oct;15(6):954–60.
34. Stafinski T, Menon D, Davis C, McCabe C. Role of centralized review processes for making reimbursement decisions on new health technologies in Europe. *Clin. Outcomes Res. CEOR*. 2011 Aug 30;3:117–86.
35. Clement FM, Harris A, Li JJ, Yong K, Lee KM, Manns BJ. Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada. *JAMA*. 2009 Oct 7;302(13):1437–43.
36. Pujolras LM, Cairns J. Why do some countries approve a cancer drug and others don't? *J. Cancer Policy*. 2015 Jun 1;4:21–5.
37. Nicod E, Kanavos P. Commonalities and differences in HTA outcomes: A comparative analysis of five countries and implications for coverage decisions. *Health Policy*. 2012 Dec;108(2–3):167–77.
38. Grepstad M, Kanavos P. A comparative analysis of coverage decisions for outpatient pharmaceuticals: Evidence from Denmark, Norway and Sweden. *Health Policy*. 2015 Feb;119(2):203–11.
39. Vegter S, Rozenbaum MH, Postema R, Tolley K, Postma MJ. Review of regulatory recommendations for orphan drug submissions in the netherlands and scotland: focus on the underlying pharmacoeconomic evaluations. *Clin. Ther.* 2010 Aug;32(9):1651–61.
40. Blankart CR, Stargardt T, Schreyögg J. Availability of and access to orphan drugs: an international comparison of pharmaceutical treatments for pulmonary arterial hypertension, Fabry disease, hereditary angioedema and chronic myeloid leukaemia. *PharmacoEconomics*. 2011;29:63–82.
41. NICE. Appraisal Guidance 341, 06/2015 [Internet]. [cited 2016 Jan 28]. Available from: <https://www.nice.org.uk/guidance/ta341>
42. NICE. Appraisal Guidance 287, 06/2013 [Internet]. Available from: <https://www.nice.org.uk/guidance/ta287/chapter/4-Consideration-of-the-evidence>
43. Cerri KH, Knapp M, Fernandez J-L. Decision making by NICE: examining the influences of evidence, process and context. *Health Econ. Policy Law*. 2014 Apr;9(2):119–41.
44. 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 Mai;13(5):437–52.
45. Dakin H, Devlin N, Feng Y, Rice N, O'Neill P, Parkin D. THE INFLUENCE OF COST-EFFECTIVENESS AND OTHER FACTORS ON NICE DECISIONS. *Health Econ.* 2014 Sep 23;

46. Mshelia I, White R, Mukku SR. An Investigation Into The Key Drivers Influencing The Decision Making Of The Scottish Medicines Consortium. *Value Health*. 2013;16(3):A264.

Tables

Table 1. Differences and Similarities of Assessment Processes in Germany, England Scotland and Australia

<i>Characteristic</i>	<i>HTA^a Agency</i>	FJC^b	NICE^c	SMC^d	PBAC^e
<i>Consideration of comparative effectiveness in final decision</i>		Yes	Yes	Yes	Yes
<i>Consideration of cost effectiveness in final decision</i>		No	Yes	Yes	Yes
<i>2-stage-process (assessment / appraisal)</i>		Yes	Yes	Yes	Yes
<i>Appraisal selection criteria</i>		All newly licensed pharmaceuticals and new indications	If mandated by the Department of Health	All newly licensed pharmaceuticals and new indications	Manufacturer submission
<i>Approx. process duration</i>		6 months	9-12 months	3 months	4-5 months

^a HTA=Health Technology Assessment; ^b FJC=Federal Joint Committee; ^c NICE=National Institute for Health and Care Excellence; ^d SMC=Scottish Medicines Consortium; ^e PBAC=Pharmaceuticals Benefits Advisory Committee

Table 2. NICE, SMC and PBAC Decisions and Comparative Effectiveness Estimates by FJC Decision

a. NICE Decisions by FJC Decisions				b. NICE Comparative Effectiveness by FJC Decisions				
NICE	FJC	Added Benefit	No Added Benefit	Sum	NICE	Added Benefit	No Added Benefit	Sum
<i>Recommended</i>		14	18	32	<i>Better</i>	19	21	40
<i>Restricted</i>		2	7	9	<i>Similar/not proven</i>	5	10	15
<i>Not recommended</i>		8	6	14	<i>Sum</i>	24	31	55
<i>Sum</i>		24	31	55				
<i>Cohen's Kappa: -0.13</i>				<i>Cohen's Kappa: 0.11</i>				

c. SMC Decisions by FJC Decisions				d. SMC Comparative Effectiveness by FJC Decisions*				
SMC	FJC	Added Benefit	No Added Benefit	Sum	SMC	Added Benefit	No Added Benefit	Sum
<i>Accepted</i>		32	48	80	<i>Better</i>	36	30	66
<i>Restricted</i>		14	20	34	<i>Similar/not proven</i>	29	71	100
<i>Not recommended</i>		19	33	52	<i>Sum</i>	65	101	166
<i>Sum</i>		65	101	166				
<i>Cohen's Kappa: 0.03</i>				<i>Cohen's Kappa: 0.26</i>				

e. PBAC Decisions by FJC Decisions				f. PBAC Comparative Effectiveness by FJC Decisions				
PBAC	FJC	Added Benefit	No Added Benefit	Sum	PBAC	Added Benefit	No Added Benefit	Sum
<i>Recommended</i>		37	45	82	<i>Better</i>	33	16	49
<i>Rejected</i>		12	21	33	<i>Similar/not proven</i>	20	50	70
<i>Deferred</i>		4	0	4	<i>Sum</i>	53	66	119
<i>Sum</i>		53	66	119				
<i>Cohen's Kappa: 0.07</i>				<i>Cohen's Kappa: 0.38</i>				

Figure

Figure 1. Early Benefit Assessments (EBAs) and corresponding appraisals of FJC, NICE, SMC and PBAC

