Link between Process and Appraisal in Coverage Decisions: An Analysis with Structural Equation Modeling

Katharina E. Fischer, PhD, Björn Stollenwerk, PhD, Wolf H. Rogowski, PhD

Background. *To achieve fair-coverage decision making, both material criteria and criteria of procedural justice have been proposed. The relationship between these is still unclear.* ***Objective***. *To analyze hypotheses underlying the assumption that more assessment, transparency, and participation have a positive impact on the reasonableness of coverage decisions.* ***Methods***. *We developed a structural equation model in which the process components were considered latent constructs and operationalized by a set of observable indicators. The dependent variable ''reasonableness'' was defined by the relevance of clinical, economic, and other ethical criteria in technology appraisal (as opposed to appraisal based on stakeholder lobbying). We conducted an Internet survey among conference participants familiar with coverage decisions of third-party payers in industrialized countries between 2006 and 2011. Partial least squares path modeling (PLS-PM) was used, which allows analyzing small sample sizes without distributional assumptions. Data on 97 coverage decisions from 15 countries and 40 experts were used for model estimation.* ***Results***. *Stakeholder participation (regression coefficient [RC] =0.289;* P *= 0.005) and scientific rigor of assessment (RC = 0.485;* P *< 0.001) had a significant influence on the construct of reasonableness. The path from transparency to reasonableness was not significant (RC = 0.289;* P *= 0.358). For the reasonableness construct, a considerable share of the variance was explained (*R^2 *= 0.44). Biases from missing data and nesting effects were assessed through sensitivity analyses.* ***Limitations***. *The results are limited by a small sample size and the overrepresentation of some decision makers.* ***Conclusions***. *Rigorous assessment and intense stakeholder participation appeared effective in promoting reasonable decision making, whereas the influence of transparency was not significant. A sound evidence base seems most important as the degree of scientific rigor of assessment had the strongest effect.* ***Key words:*** *formulary decision making; pharmaceuticals; pharmacist; statistical methods; survey methods; formulary decision making.* ***(Med Decis Making 2013;33:1009–1025)***

Faced with escalating health care costs, various third-party payers across industrialized countries have established processes to decide on the coverage of health technologies.[1] In contrast to the market authorization of new drugs, coverage decision processes and appraisal criteria are highly heterogeneous.[2,3] As a consequence, decision outcomes differ across committees.[4] Understanding variations in both criteria and processes is equally relevant to patients, industry, clinicians, government, and the public because they determine the availability of technologies beyond market authorization.

Conceptual frameworks describing key elements of coverage decision processes typically distinguish between assessment and appraisal.[5–8] In the

Received 17 February 2012 from Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Health Economics and Health Care Management, München, Germany (KEF, BS, WHR); Universität Hamburg, Hamburg Center for Health Economics, Hamburg, Germany (KEF); and Ludwig-Maximilians-University, Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Clinical Center, München, Germany (WHR). This research was carried out on behalf of Helmholtz Zentrum München–German Research Center for Environmental Health (HMGU). The HMGU is an independent organization funded by the German and Bavarian governments. Revision accepted for publication 7 March 2013.

DOI: 10.1177/0272989X13490837

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://mdm.sagepub.com/supplemental.

Address correspondence to Katharina E. Fischer, PhD, Universität Hamburg, Hamburg Center for Health Economics, Esplanade 36, 20354 Hamburg, Germany; telephone: + 49-89-42838-9296; fax: + 49-89-42838-8043; e-mail: katharina.fischer@wiso.uni-hamburg.de.

assessment step, potential outcomes are analyzed to extract information about a product's effectiveness and costs or cost-effectiveness. In the appraisal step, the desirability of assessed outcomes is judged through definition of criteria, which are based on the values of the individual or committee that decides on the technology. Further elements that are frequently incorporated in conceptual frameworks are the exchange of information and stakeholder participation.[2,4,9] Given the complexity of decision processes, a large number of further process elements can be included in the analysis, as has been done by Hutton and others.[10] Given the time constraints of respondents, surveys have to restrict the assessed items to a small number that can be justified theoretically and surveyed empirically.

From the viewpoint of ethical theory, both material and procedural criteria have been proposed to guide legitimate decision making about scarce health care resources. The material criteria typically include a technology's clinical effectiveness, economic criteria such as cost-effectiveness, and other ethical criteria such as severity of disease or equal access.[11,12] Therefore, both the criteria relevant in appraisal, as well as the generation of evidence about to what extent a technology meets the criteria of effectiveness and cost-effectiveness, appear to be important in the analysis of coverage decision processes from an ethical perspective.

A frequently cited procedural framework is ''accountability for reasonableness'' by Daniels and Sabin.[13] Given that reasonable people may disagree about which material criteria should determine the decision outcome, they only postulate that the decision should be based on evidence and appraisal criteria, which ''all fair-minded parties (managers, clinicians, patients, and consumers in general) can agree are relevant to deciding how to meet the diverse needs of a covered population under necessary resource constraints.''[13] As a complement, they propose transparency and participation in the minimum form of a ''mechanism for challenge and dispute resolution . . ., including the opportunity for revising decisions in light of further evidence or arguments.''[13] The framework has been criticized for demanding too little participation to achieve truly legitimate decision outcomes in the absence of a specification of material criteria.[14] The question of how much transparency or participation is desirable is thus an unresolved issue.[15]

A recent economic theory of the fourth hurdle provides a theoretical framework to address assessment, appraisal, transparency, and participation in one common framework.[9] It interprets procedural criteria as a means to promote material ones: transparency and participation can be seen as investments to ensure that decisions are indeed oriented at the value judgments of the covered population rather than on the opportunistic preferences of their decision-making agents. These investments are justified on economic grounds if the expected costs of increased transparency and participation fall below the expected costs associated with the agency problems. Investments in evidence generation can be judged by similar criteria[16]: here, the value of reducing decision uncertainty can be compared with the costs of evidence generation.[9]

Both the considerations from ethical and economic theory are based on the assumption that more assessment, transparency, and participation have an impact on decision making and that this association is considered positive. Given the complexity of decision problems and the values that are reflected in each decision,[6] neither the ethical nor the economic frameworks provide a standardized characterization of ''good'' decisions. Instead, they refer to reasonable medical, economic, and other ethical criteria. It is empirically unclear, though, to what extent this is indeed the case.

Empirical research has addressed various aspects of coverage decision processes but barely incorporated theoretical considerations.[17] Studies that analyzed real-world decisions focused on influences on decision outcomes by cost-effectiveness considerations and the quality of the available evidence, analysis of single criteria, and committees and selected process characteristics such as the timing after market authorization.

The objective of our study is to extend this work and analyze hypotheses on the relation between process components and the use of reasonable appraisal criteria in decision making on health technologies. We state a structural equation model (SEM), which we estimated with partial least squares path modeling as the modeling technique using data on real-world decisions from industrialized countries.

METHODS

Definition of Concepts and Hypotheses

Our hypotheses are based on the following 4 components that have been discussed in the context of evidence-based decision making, as well as procedural and distributive justice.

- **Transparency (η_1):** Processes are considered transparent if relevant information is provided so that

decisions can be retraced.[18,19] More transparency improves the extent to which a decision can be controlled. Transparency is reflected by the volume of documentation available for a decision and the degree of detail in the documentation of processes and decision outcomes.[18,20,21]

- **Participation (η_2):** Participation implies that different stakeholder groups are involved at various stages of decision processes to ensure that their interests are not neglected. Stakeholder groups may include manufacturers, patients, or government or service providers. Involvement may be in the form of providing information, commenting, appealing, or voting.[20,22–25]
- **Scientific rigor (of assessment) (η_3):** The assessment stage has been characterized by methods that ground on evidence-based medicine, comparative effectiveness research, and health technology assessment (HTA) to make statements about the technology's effectiveness and cost-effectiveness.[26,27] The information that is produced depends on the development stage of the product and the assessment requirements stated by decision makers.[4,20] Assessments depend on scientific judgments about the exact evaluation methods used for a technology. Thus, scientific rigor may vary across decisions and committees due to variation in methods and level of evidence. We defined scientific rigor by the methodological standards for generating evidence. The assessment of effectiveness may range from collecting expert opinions to quantitative meta-analyses of studies. Assessment of costs may go from rough estimates to comprehensive cost-effectiveness or budget impact analyses. Rigorous assessments are prerequisites to reasonable decisions that are evidence based and accepted by informed people.[18,19]
- **Reasonableness (η_4):** Technology appraisal involves value judgments about the technology. The criteria used in technology appraisal have been addressed by a large body of theoretical and empirical research.[11,12,28–33] They can be selected implicitly, ad hoc, history based, or as a result of priority-setting exercises.[2,4,28,29] It is frequently argued that in a pluralistic society, there is not one single criterion but a set of criteria.[11,12] Thus, reasonableness is difficult to operationalize because what is considered reasonable depends on potentially varying value judgments of decision makers and stakeholders as well as the situation. In this model, we propose to use the extent to which typically accepted criteria are considered in technology appraisal: the use of clinical, economic, and other ethical considerations (as opposed to the role of lobby interests of stakeholders).[2,30,34] Although we neither weigh the criteria nor predefine a certain combination, we assume that the higher the relevance of each criterion considered, the stronger the degree of reasonableness is reflected.

We considered the components participation, transparency, and scientific rigor of assessment as process components that ultimately influence the reasonableness of appraisal. Accordingly, we stated 5 hypotheses:

Hypothesis 1 (transparency ⟶ reasonableness): The higher the degree of transparency, the higher the degree of reasonableness (in terms of the relevance of clinical, economic, and other ethical) decision-making criteria because this facilitates a better control of the decision makers.[9]

Hypothesis 2 (participation ⟶ reasonableness): The higher the degree of participation, the higher the degree of reasonableness because the higher the degree of participation, the better stakeholders and the covered population can ensure that the technology is appraised against criteria they consider reasonable.[9]

Hypothesis 3 (scientific rigor ⟶ reasonableness): The higher the degree of scientific rigor, the higher the degree of reasonableness because decision makers can draw upon better evidence regarding whether the criteria are met.[9]

Hypothesis 4 (transparency ⟶ scientific rigor): The higher the degree of transparency, the higher the degree of scientific rigor of assessment because the methodological quality can be better controlled by the scientific community.

Hypothesis 5 (participation ⟶ scientific rigor): The higher the degree of stakeholder participation, the higher the degree of scientific rigor of assessment. If more stakeholders participate in different stages of decision making, more evidence is identified and improvements of a weak evidence basis can be enforced more easily.

Development of a Structural Equation Model

To analyze the relationships between the components, an empirical method is needed that has several requirements. The components cannot be measured directly. For example, transparency is a concept that cannot be observed from documentation but can be considered as a latent construct. The hypotheses suggested multiple and interrelated dependence relationships. Thus, the method should capture the set of linkages between process components.

SEM techniques—a combination of factor and multiple regression analysis—are capable of measuring interrelationships between latent constructs.[35] SEM inherits several properties that differ from multivariate regression, which seem beneficial in our context.[36] In regression analysis, one dependent variable is typically explained by several independent

variables. Multiple but nonhierarchical endogenous constructs (here: reasonableness and scientific rigor) would require specification of several regression models. Using SEM, this can be performed in 1 step. Multivariate regression does not easily allow the specification of the relationship between a latent construct and its observable indicators. Instead, the observable indicators that describe a latent construct would need to be combined into an aggregate score that may over- or underestimate the influence of single indicators. In SEM, the relevance of each indicator is determined in the model so that no bias occurs from arbitrary definition of latent construct scales. In addition, 1 branch of SEM offers estimation techniques that do not require distributional assumptions that we could not state reliably for most indicators.

We defined an SEM considering each component as a latent construct described by a set of observable indicators. For operationalization, we developed an Internet survey based on the framework of Rogowski and others[37] and a corresponding structured survey instrument for data collection.[20] The framework describes the stylized steps of a process from the point where a technology enters a health care market to diffusion into routine use. One question was stated per indicator in an online questionnaire (available from the authors upon demand). The SEM combines the linkages between constructs in the structural model and the relations between constructs and indicators in the measurement models (see Figure 1). All constructs were defined in the reflective mode, meaning that the causality is directed from the construct to the indicator. Table 1 describes all constructs and indicators.

Formally, the structural model can be written as

$$\begin{bmatrix} \eta_3 \\ \eta_4 \end{bmatrix} = \begin{bmatrix} \beta_{11} & \beta_{22} & 0 \\ \beta_{12} & \beta_{21} & \beta_{31} \end{bmatrix} \cdot \begin{bmatrix} \eta_1 \\ \eta_3 \\ \eta_4 \end{bmatrix} + \begin{bmatrix} \zeta_3 \\ \zeta_4 \end{bmatrix}, \qquad (1)$$

where β_{11} and β_{22} denote the path coefficients of the 2 exogenous constructs, transparency η_1 and participation η_2, on the construct scientific rigor; β_{12}, β_{21}, and β_{31} are the path coefficients of the constructs transparency, participation, and scientific rigor to the construct reasonableness η_3; and ξ_3 and ξ_4 represent the endogenous constructs' residuals. In the measurement models of equation (2), each indicator y_{ij} of a construct η_i, $i = 1, \ldots, 4$ is a linear function of its loading λ_{ij} and the residual ε_{ij},. The index j denotes the indicator y_j of the construct η_i, $j = 1, \ldots, k_i$.

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \\ y_{31} \\ y_{32} \\ y_{41} \\ y_{42} \\ y_{43} \\ y_{44} \\ y_{45} \\ y_{46} \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 & 0 & 0 \\ \lambda_{12} & 0 & 0 & 0 \\ 0 & \lambda_{21} & 0 & 0 \\ 0 & \lambda_{22} & 0 & 0 \\ 0 & \lambda_{23} & 0 & 0 \\ 0 & \lambda_{24} & 0 & 0 \\ 0 & 0 & \lambda_{31} & 0 \\ 0 & 0 & \lambda_{32} & 0 \\ 0 & 0 & 0 & \lambda_{41} \\ 0 & 0 & 0 & \lambda_{42} \\ 0 & 0 & 0 & \lambda_{43} \\ 0 & 0 & 0 & \lambda_{44} \\ 0 & 0 & 0 & \lambda_{45} \\ 0 & 0 & 0 & \lambda_{46} \end{bmatrix} \cdot \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{41} \\ \varepsilon_{42} \\ \varepsilon_{43} \\ \varepsilon_{44} \\ \varepsilon_{45} \\ \varepsilon_{46} \end{bmatrix}. \qquad (2)$$

Survey Instrument and Data

We conducted a retrospective cross-sectional survey on decisions made on health technologies in European Union (EU) member states and Organization for Economic Co-operation and Development (OECD) countries between 2006 and 2011.

Sampling of Observation Units

A decision was defined as an appraisal procedure to determine reimbursement of a primary third-party payer. We equally considered advisory and decision-making committees. While decision-making committees are in charge of making legally binding decisions, advisory committees make recommendations to a higher authority rather than decisions.[38] However, recommendations are frequently adopted without challenge. Each decision was identified by country, payer, deciding committee, technology, health condition, and further specifications (e.g., restrictions to patient subgroups). If decisions were made at a regional level, the instrument captured the geographical area.

Identification of Respondents

Experts within the field of HTA and economic evaluation who had observed or participated in decision processes were identified from the abstract books of 2 conferences: the HTA International (HTAi) meeting in 2010 in Dublin, Ireland,[39] and the congress of the International Health Economics Association in 2011 in Toronto, Canada.[40] Email addresses were identified via the membership directories, Google, and the PubMed database. Abstracts were screened by title and needed to relate to coverage, HTA, or economic evaluation. Participants' affiliations or abstract titles should indicate

Table 1 Specification of Measurement Models for Structural Equation Model of Coverage Decision Making.

Construct	Description	Indicator(s)	Indicator description	Operationalization
Participation	Degree of stakeholder involvement during decision process	Number of different types of participating stakeholders	Number of types of stakeholders involved in decision process. High diversity of stakeholders increases the possibility that particular interests of single stakeholders are balanced out.	Types of stakeholders, equally weighed: Service provider(s) Payer Government Patients/patient representative(s) Industry
		Degree of stakeholder involvement	Number of types of stakeholders involved at stages in decision process. More involvement opportunities result in stronger participation.	One indicator for each type of involvement: information provision, appeal stated during or after decision, voting, stakeholders equally weighed
Transparency	Volume and degree of detail of documentation available in terms of published information	Volume of information published during or after decision process	Degree of transparency reflected by the volume of documents published for each decision process	Types of documents, equally weighted: Outcome-related documents: decision outcome, decision rationale, public appraisal meeting, or minutes published Process-related documents: rationale for choosing decision object, health technology assessment (HTA) report, stakeholder comments
		Type of information provided	Degree of transparency reflected by the diversity of published information provided (i.e., whether it relates to the process or decision outcome or both)	No information available Only process-related information available Only outcome-related information available Outcome- and process-related information available Full documentation

(continued)

Table 1 (continued)

Construct	Description	Indicator(s)	Indicator description	Operationalization
Scientific rigor of assessment	Degree of methodological standard applied for evidence generation	Scientific rigor in assessment of effectiveness	The degree of scientific rigor is positively reflected by the degree of methodological standards used for the assessment of effectiveness.	No assessment of effectiveness At least based on expert opinion At least based on consideration of published literature At least based on a systematic literature review/HTA At least based on original quantitative meta-analysis Assessment of effectiveness based on own study
		Scientific rigor in assessment of costs/ cost-effectiveness	The degree of scientific rigor is positively reflected by the degree of methodological standards used for the assessment of costs/ cost-effectiveness (CE).	No assessment of costs/CE At least based on a review of economic studies At least based on other cost analysis Based on scientific budget impact, cost-effectiveness, or cost-utility analysis
Reasonableness	Extent to which decision can be considered reasonable	Relevance of criteria that contribute to reasonable appraisal	The higher the relevance of clinical, economic, or other ethical criteria, the higher the degree of reasonableness of the decision.	Relevance of each criterion (not relevant/relevant/strongly relevant); one indicator per criterion: Clinical: effectiveness: health benefit; effectiveness: other benefit (e.g., knowledge of diagnostic test result) Economic: cost-effectiveness, budget impact Other ethical criteria: severity of the disease, equitable access to health care

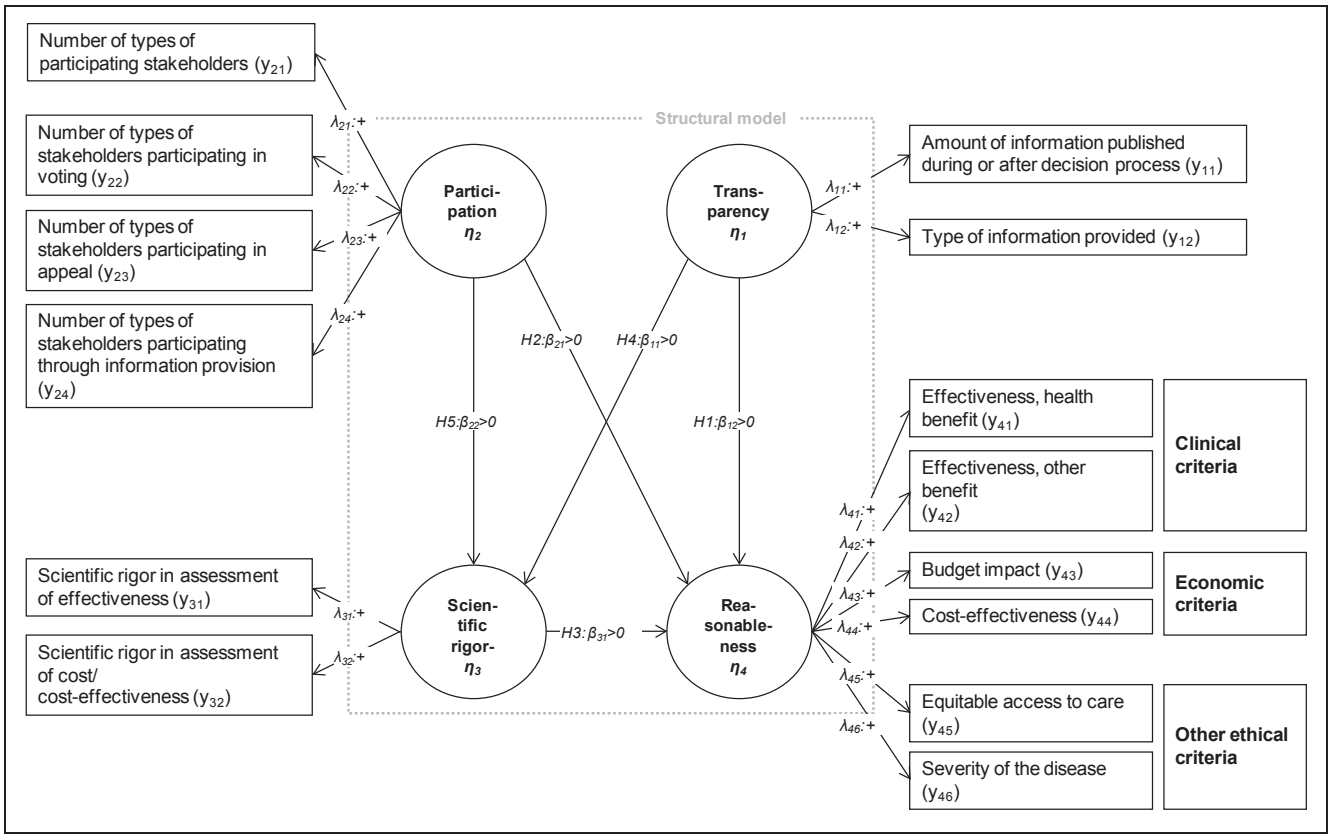

Figure 1 Structural equation model for coverage decision making.

that experts could provide information for EU or OECD member countries. To obtain a balanced sample in terms of the countries represented, we applied a form of quota sampling. We distributed countries with more than 5 eligible abstracts into 2 waves, each with 5 respondents, as some were overrepresented because of hosting the congress (Ireland, Canada), having larger populations (United States), or having proceeded faster in the development of economic evaluation (United Kingdom). In total, 279 individuals were selected.

Internet Survey

Between June and October 2011, we invited respondents via email to participate in our Internet survey. It was implemented with EFS Survey Version 8.0 (Unipark GmbH, Hürth, Germany). To limit the survey length, we restricted the number of decisions to 4. To minimize recall bias and to focus on recent decision making, we asked respondents for decisions made within the past 5 years. The invitation related to the abstract and contained information on the survey background. To increase survey uptake, nonresponders were reminded at least twice. Respondents could provide contact details of individuals whom they considered eligible for the survey. Accordingly, we invited 31 additional respondents referred from the initial contact. The survey procedure was piloted in May 2011. We assessed the comprehensibility of the questions, the technical implementation, and the procedure for contacting respondents among 9 experts of the EU project HIScreenDiag. To test for appropriate reaction of conference participants, we invited 12 HTAi participants not included in the main mailing. No ethics approval was requested as no patient-related data were collected. No external funding was obtained for conducting the survey.

Data

The survey retrieved 119 decisions from 40 experts in 15 countries, of which 97 observations were used for model estimation. The response rate was 23%. Another 18% of respondents replied that they did not know decisions or referred to other experts. For verifying each decision specified by respondents,

we collected available documentation on the Internet. If respondents did not provide information, we searched the payers' or decision makers' Web sites. Respondents were contacted in case of ambiguities regarding the scope of the decision. To categorize decision objects, we assigned each technology to a unique health care system function according to the World Health Organization (WHO)/OECD System of Health Accounts 2011.[41] Health conditions were categorized by the *International Statistical Classification of Diseases and Related Health Problems Version 10* (*ICD-10*).[42] Twenty-five decisions could not be validated by corresponding decision documentation but were kept for data analysis. Data from the pilot study were kept as the questionnaire was not changed. We transformed all questions containing multiple-item answers into ordinal scales (see Table 2 for a definition). Data preparation was performed with SAS version 9.2 (SAS Institute, Cary, NC).

We had to apply an algorithm to replace missing data because some respondents gave partial information. Four respondents did not specify a decision, and another 4 respondents provided data that only specified the decision object but no information about the third-party payer and deciding committee. Thus, the decision was not attributable. Another 7 observations were excluded in which more than 60% of questions were not answered. For 12 observations with missing values, we applied conditional mean imputation. We imputed the mean for the same decision-making institution or, if not represented, the country. If no data could be imputed, which was the case for 4 observations, SmartPLS (Hamburg University of Technology, Institute for Human Resource Management and Organizations, Hamburg, Germany) substituted the overall sample mean. Missing values were more prevalent at the end of the questionnaire. The summary statistics (Table 2) include information on missing values per indicator.

Model Estimation and Evaluation

Two types of modeling techniques are available for SEM that depend on the optimization algorithm.[43] In partial least squares path modeling (PLS-PM), the share of the variance that is explained for the endogenous constructs by the SEM is maximized through a series of ordinary least squares regressions. In the covariance-based techniques (LISREL, AMOS), maximum likelihood estimators extract a set of model parameters so that the theoretical covariance matrix is as close as possible to the empirical covariance matrix.

We selected PLS-PM for several reasons: contrary to the covariance-based techniques of SEM, it allows model estimations with small sample sizes and without making distributional assumptions.[44,45] It is suitable for exploration and prediction of effects at early stages in theoretical development. In contrast, covariance-based techniques are better suited for confirmatory analysis but require knowledge of parameter distributions and larger sample sizes.[43] If the sample size is large, covariance-based methods are preferable because of better parameter consistency and accuracy. If the focus is on prediction and theory development—as in our case—PLS-PM should be preferred.[43] For the underlying SEM, a feasibility test to assess whether plausible estimation measures can be obtained from the PLS-PM algorithm has been performed using data on decisions of newborn screening technologies.[46] For estimation, we used the SmartPLS version 2.0.[47] Through application of an iterative procedure, the PLS-PM algorithm calculates the path coefficients between the latent constructs and the scores of the constructs in the structural model and the weights and loadings of the manifest indicators in the measurement models in a sequence of ordinary least squares regressions.[44] The path weighting scheme was applied to calculate the inner weights in the iteration procedure.

We assessed the model estimation against established reliability and validity measures for PLS-PM at the level of the measurement and structural model.[36,44,48] To obtain asymptotic t statistics, we applied a bootstrapping procedure with 5000 samples. In contrast to covariance-based approaches and regression analysis, assessing the model's overall goodness of fit is limited because a distribution-free variance is assumed.[48]

Sensitivity Analyses

We performed sensitivity analyses to assess influences from missing data, nonresponse bias, and nesting effects. Regarding missing data, we reestimated the model with the 85 complete observations. However, we used the imputed data set as the base case. Furthermore, we compared the data set with complete cases only with the imputed data set for nonrandom influences from the type of technology (pharmaceuticals v. others) and involvement of respondents in the decision (direct v. indirect) with the Fisher exact test. Regarding nonresponse bias, we assessed whether the proportion of survey respondents differed from the individuals who were contacted but did not respond to the survey by country. To analyze whether data were missing at random regarding the scope of decision, we further compared

Table 2 Summary Statistics of Indicators.

Construct/Indicator and Categories	Observations	Mean (SD) or No. (%)
Construct: Participation, mean (SD)		
Number of stakeholders involved in decision process	97	2.84 (1.23)
Number of stakeholders participating through information provision	97	1.31 (1.10)
Number of stakeholders participating in appeal	97	0.31 (0.60)
Number of stakeholders participating in voting	97	1.13 (0.98)
Construct: Transparency		
Type of information provided, No. (%)		
Missing		3 (3.09)
0—No information available		6 (6.19)
1—Only process-related information available		13 (13.4)
2—Only outcome-related information available		17 (17.53)
3—Outcome- and process-related information available		51 (52.58)
4—Full documentation		7 (7.22)
Number of documents reported, mean (SD)	94	2.05 (1.39)
Construct: Scientific rigor of assessment, No. (%)		
Scientific rigor in assessment of effectiveness		
Missing		4 (4.12)
0—No assessment of effectiveness		2 (2.06)
1—At least based on expert opinion		4 (4.12)
2—At least based on consideration of published literature		8 (8.25)
3—At least based on a systematic literature review/health technology assessment		28 (28.87)
4—At least based on original quantitative meta-analysis		13 (13.4)
5—Assessment of effectiveness based on own study		38 (39.18)
Scientific rigor in assessment of costs/cost-effectiveness		
Missing		9 (9.28)
0—No assessment of costs/cost-effectiveness		7 (7.22)
1—At least based on a review of economic studies		7 (7.22)
2—At least based on other cost analysis		12 (12.37)
3—Based on scientific budget impact, cost-effectiveness, or cost-utility analysis		62 (63.92)
Construct: Reasonableness, No. (%)		
Effectiveness, health gain		
Missing		8 (8.25)
0—Not relevant		14 (14.43)
1—Relevant		15 (15.46)
2—Strongly relevant		60 (61.86)
Effectiveness, other benefit		
Missing		8 (8.25)
0—Not relevant		60 (61.86)
1—Relevant		21 (21.65)
2—Strongly relevant		8 (8.25)
Budget impact		
Missing		8 (8.25)
0—Not relevant		41 (42.27)
1—Relevant		23 (23.71)
2—Strongly relevant		25 (25.77)
Cost-effectiveness		
Missing		8 (8.25)
0—Not relevant		35 (36.08)
1—Relevant		19 (19.59)
2—Strongly relevant		35 (36.08)
Effect on equitable access to health care		
Missing		8 (8.25)
0—Not relevant		51 (52.58)

(continued)

Table 2 (continued)

Construct/Indicator and Categories	Observations	Mean (SD) or No. (%)
1—Relevant		27 (27.84)
2—Strongly relevant		11 (11.34)
Severity of the disease		
Missing		8 (8.25)
0—Not relevant		40 (41.24)
1—Relevant		28 (28.87)
2—Strongly relevant		21 (21.65)

qualitatively whether the decisions that were excluded because of missing values differed by health condition and technology from those included in the data set. To account for potential nesting effects from the clustering of decisions by respondents, we randomly drew a reduced data set with 31 complete observations in which each respondent was represented just once. We compared reliability and validity of the SEM with the estimation of the full data set.

RESULTS

Descriptive Statistics

Table 3 provides an overview of decisions by country, decision-making institution, and outcome. Spanish health authorities (23), the UK National Institute of Health and Clinical Excellence (NICE) (12), the German Federal Joint Committee (FJC) (10), and the Irish Health Service Executive (9) were most frequent. Although acceptance rates varied across institutions, half the decisions resulted in unrestricted coverage. The other decisions resulted about equally in restricted or no coverage.

Table 4 provides an overview of the technologies and health conditions. Most decisions were made on prescribed medicines (47), specialized inpatient curative care (16), or therapeutic appliances (9). Health conditions covered a wide range of the *ICD-10*. Some indications were more prevalent: cancer (25), diseases of the circulatory system (16), and endocrine, nutritional, and metabolic diseases (10). Decisions covered different indications except for 7 responses on vaccines against human papillomavirus and a few technologies with 2 responses.

Summary statistics of indicators are presented in Table 2. Clustering results by decision committees showed that processes were heterogeneous not only between countries but also for the same decision committees. Regarding nonresponse biases, we found that compared with our initial set of conference participants, some countries were overrepresented (Spain, Ireland, and United Kingdom), whereas others were underrepresented (Canada and United States) in the study sample. From the 7 observations that had to be excluded because of too many missing values, 5 decisions were on technologies other than pharmaceuticals.

Structural Equation Model Results

In compliance with the PLS-PM literature, we first provide reliability and validity measures for the measurement models before we present the results of the structural model.[36] The online appendix provides the values for all evaluation measures. Estimated path coefficients and indicator loadings are depicted in Figure 2.

Measurement Models

Overall, the measurement model estimations were acceptable in terms of reliability and validity at the construct and indicator levels. Composite reliability is a measure between zero and 1 for reliability at the construct level. Accounting for nonequal indicator loadings, it assesses whether indicators consistently represent the same construct. Values for composite reliability for the constructs participation and transparency were greater than the critical value of 0.7 and at least 0.66 for the constructs reasonableness and scientific rigor. A value of 0.6 is acceptable in the early research stages.[48] For reliability at the indicator level, the indicators' variance that was explained by the construct varied. This information was obtained from the indicator loadings, which denote the correlation with the construct. For the constructs participation, transparency, and scientific rigor, indicators were highly correlated with corresponding constructs except for the indicator ''number of types of stakeholders participating in voting.''

Table 3 Decisions by Country, Decision-Making Institution, and Decision Outcome.

Country	Decision-Making Institution	No. of Decisions	Decision Outcome		
			Positive, Full Coverage	Positive, Restricted Coverage	Negative
Australia	PBAC	6	1	3	2
Austria	HVB	1	1	0	0
	National Health Care Commission	1	1	0	0
Belgium	RIZIV/INAMI	6	3	2	1
Finland	MUMM program	4	2	2	0
France	HAS/CNEDIMTS	1	1	0	0
Germany	FJC	10	5	0	5
	InEK	1	0	1	0
Ireland	HSE	9	7	0	2
Italy	Regional government	2	2	0	0
Netherlands	CVZ	5	2	2	1
	Gezondheidsraad	1	0	1	0
Poland	MoH/AOTM	4	1	2	1
Spain	Basque government	4	2	1	1
	Basque Health Service	1	1	0	0
	MoH Catalonia	4	4	0	0
	MoH Galicia	4	2	1	1
	MoH La Rioja	4	1	3	0
	MoH Canary Islands	4	1	2	1
	National MoH: Interterritorial Board	2	1	1	0
Sweden	TLV	3	1	0	2
Switzerland	EDI	1	1	0	0
United Kingdom	JCVI	1	1	0	0
	NICE	12	9[a]	0	3
United States	Intermountain Healthcare Coverage Committee	4	0	2	2
	Washington State HCA	2	0	1	1

AOTM, Agency for Health Technology Assessment (Agencja Oceny Technologii Medycznych); CVZ, Health Care Insurance Board (College voor Zorgverzekeringen); EDI, Federal Department of Home Affairs (Eidgenössisches Departement des Inneren); FJC, Federal Joint Committee (Gemeinsamer Bundesausschuss); HAS/CNEDIMTS, High Health Authority, National Committee for the Assessment of Medical Devices and Health Technologies (Haute Autorité de Santé, Commission nationale d'Evaluation des Dispositifs médicaux et des Technologies de Santé); HSE, Health Service Executive; HVB, Main Association of Austrian Social Security Organisations (Hauptverband der österreichischen Sozialversicherungsträger); InEK, Institute for the Hospital Remuneration System (Institut für das Entgeltsystem im Krankenhaus); JCVI, Joint Committee on Vaccination and Immunization; MoH, Ministry of Health; MUMM, Advisory Board of the Office of HTA and 20 Finnish hospital districts; NICE, National Institute for Health and Clinical Excellence; PBAC, Pharmaceutical Benefits Advisory Committee; RIZIV/INAMI, National Institute for Health and Disability Insurance; TLV, Pharmaceutical Benefits Board (Tandvårds–och läkemedelsförmånsverket).

[a]Outcome of decision partly complemented by own validation with decision documents

Its value was below zero, which we did not hypothesize, and its absolute value was small, indicating a lack of indicator reliability. In the reasonableness construct, the indicators' correlations with the construct varied. The indicator that describes the relevance of cost-effectiveness as a decision criterion was correlated highest. Some 74% of the variance was explained by this construct. The smallest correlation was obtained for the criterion ''severity of the disease'' ($\lambda_{46} = 0.28$).

Validity is assessed by convergent and discriminant validity. Convergent validity describes whether the set of indicators uniquely represents the underlying construct. The average variance extracted (AVE) measures the variance of the indicators of the reflective construct relative to the total amount of variance, including the measurement error's variance. The AVE was greater than the suggested threshold of 0.5 for the constructs transparency ($AVE(\eta_1) = 0.9$), participation ($AVE(\eta_2) = 0.53$), and scientific rigor ($AVE(\eta_3) = 0.5$) but not for the reasonableness construct ($AVE(\eta_4) = 0.27$). Two measures were used to evaluate discriminant validity. They appraise whether the constructs are sufficiently distinct from each other. The Fornell-Larcker criterion provides a comparison of the AVE of a latent variable with the squared correlations between the construct and any other construct. For all constructs, the AVE was

Table 4 Overview of Technologies and Health Conditions.

Technology or Health Condition	No. (%)
Technology (System of Health Accounts [2.0 2011])	
HC 5.1.1 Prescribed medicines	47 (48.45)
HC 1.1.2 Specialized inpatient curative care	16 (16.49)
HC 5.2 Therapeutic appliances and other durable medical goods	9 (9.28)
HC 6.2 Immunization programs	8 (8.25)
HC 4.1 Laboratory services	7 (7.22)
HC 4.2 Imaging services	4 (4.12)
HC 6.3 Early disease detection programs	4 (4.12)
HC 5.2.3 Other orthopedic appliances, orthesis, and prosthetics (excluding glasses and hearing aids)	2 (2.06)
Health condition: *ICD-10* (Version 2007) chapters	
Certain infectious and parasitic diseases	9 (9.28)
Neoplasms	25 (25.80)
Endocrine, nutritional, and metabolic diseases	10 (10.30)
Mental and behavioral disorders	4 (4.12)
Diseases of the nervous system	3 (3.09)
Diseases of the ear and mastoid process	1 (1.03)
Diseases of the circulatory system	16 (16.5)
Diseases of the respiratory system	1 (1.03)
Diseases of the digestive system	7 (7.21)
Diseases of the musculoskeletal system and connective tissue	8 (8.24)
Diseases of the genitourinary system	3 (3.09)
Pregnancy, childbirth, and the puerperium	1 (1.03)
Certain conditions originating in the perinatal period	1 (1.03)
Symptoms, signs, and abnormal clinical and laboratory findings, not classified elsewhere	1 (1.03)
Injury, poisoning, and certain other consequences of external causes	1 (1.03)
Cannot be attributed	5 (5.15)

ICD-10, International Statistical Classification of Diseases and Related Health Problems Version 10.

larger than any other squared correlation, suggesting that the indicator sets are sufficiently related to the construct. Second, the indicators' cross-loadings with other constructs were compared. These were at maximum at the assigned construct for all indicators except for the ''number of stakeholders participating in voting'' (y_{22}) of the participation construct.

Structural Model

At structural model level, 3 of the 5 paths had the expected sign and were significant. The degree of reasonableness was significantly influenced by the degree of participation and scientific rigor of assessment, which supported hypotheses 2 and 3. Their path coefficients, which can be interpreted as standardized β coefficients, were significant with P values smaller than 0.001. The effect of scientific rigor on the degree of reasonableness was nearly twice as large ($\beta_{31} = 0.49$) as that of participation ($\beta_{21} = 0.29$). Furthermore, the results supported hypothesis 5 as the scientific rigor was significantly influenced by the degree of participation ($P < 0.001$). Hypotheses 1 and 4 were not supported as the degree of transparency had a nonsignificant effect on the endogenous constructs.

The determination coefficient R^2—analogous to multiple regression—reflected that the share of the variance explained was 44.3% for the construct reasonableness and 13.8% for the construct scientific rigor. The goodness-of-fit index, which is defined here by the geometric mean of the average communality and the model's average R^2, was 0.37.[49] It measures the average of all the squared correlations between each indicator and the corresponding latent construct. We consider this value sufficiently high because of the high contribution of R^2 of the reasonableness construct and the value of the geometric mean of the communality (0.468).[45] However, the goodness of fit of PLS models has to be handled with caution because it includes R^2 for which acceptable values depend on the research context, and thus no single consented threshold exists.[48] The value of the average redundancy that measures the portion of the variability of the indicators connected to an endogenous construct explained by the latent construct directly connected to the block was 0.057.[45]

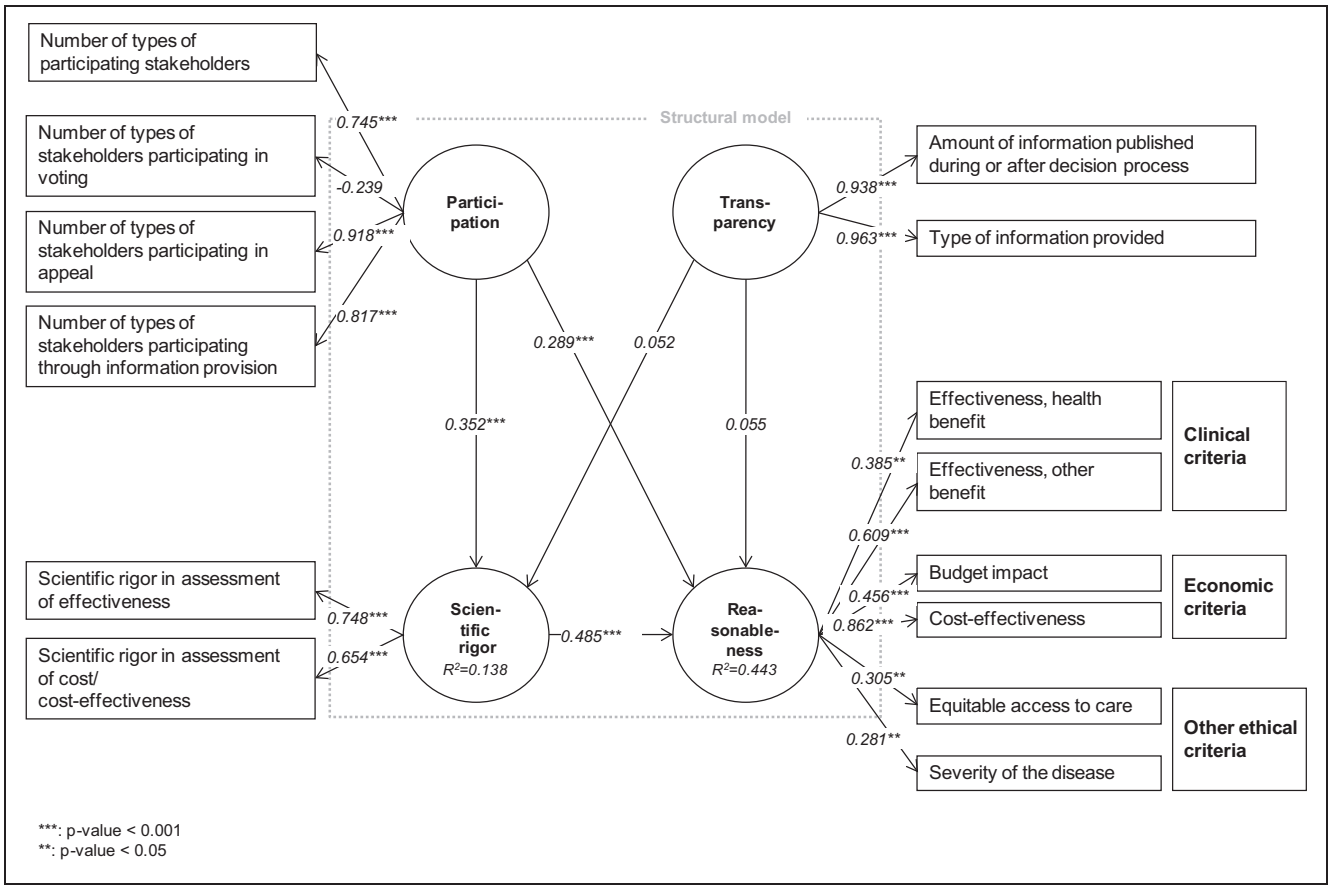

Figure 2 Structural equation model: estimation results.

Measures of the effect size f^2 revealed that the exogenous construct participation η_2 weakly contributed to the explanation of both endogenous constructs (η_3, η_4). F^2 values for the construct transparency indicated that it did not contribute to an explanation of the endogenous constructs.

Sensitivity analyses showed that the model results did not change fundamentally if the data set was restricted to the 85 observations without missing values and if only 1 observation from each respondent was used to avoid nesting effects. Furthermore, no significant difference between observations regarding missing values was identified depending on the type of technology and respondents' involvement decisions.

DISCUSSION

In both the ethical and the economic literature, theoretical considerations about decision processes have been based on the assumption that more assessment, transparency, and participation have a positive impact on the reasonableness of decision criteria. This is the first study that investigates this assumption with SEM methods.

Five hypotheses on the relation of the constructs transparency, participation, and scientific rigor on one another and on the construct reasonableness were specified, of which 3 were significant. According to the model, the degree of stakeholder participation and the scientific rigor of assessment positively influence the decisions' degree of reasonableness. The effect of the amount of released information was not statistically significant.

The scientific rigor of assessment had the highest positive association with our construct of reasonable decision making. This indicates that a sound evidence base may not only disclose relevant data on a health technology but also promote the application of reasonable appraisal criteria. Although this has been a frequently stated claim, our estimation confirms this notion.[50]

Stakeholder participation displayed the second highest positive association with the degree of reasonableness. Strong stakeholder involvement may facilitate that singular interests are balanced and criteria that can be weighed against existing evidence play a higher role in decision making. This empirical result strengthens Friedman's criticism of the ''accountability for reasonableness'' framework that more participation than just a mechanism of appeal would be desirable to achieve legitimate and fair decisions about health care resources.[14]

The positive association of the degree of reasonableness and the 2 components of scientific rigor and participation can also be seen as a first empirical confirmation of relations assumed in the ''accountability for reasonableness'' framework[13] as well as the economic theory of the fourth hurdle.[9]

Our results further suggest that transparency matters less for making reasonable decisions, particularly if there is a sound scientific evidence base and sufficient possibilities for participation. Even if promoting transparency is a frequent claim in discussions about decision making, our estimation suggests that—contrary to our hypotheses—transparency may not be influential for achieving high levels of scientific rigor and for promoting reasonableness. Nevertheless, a high degree of transparency is likely to be beneficial for other reasons: to ensure consistency and legality of value judgments, to inform involved stakeholders about the coverage status of the technology, and to foster possibilities for participation.[50,51] This corresponds with the role of transparency in rulings of social courts: according to Syrrett,[52] the primary function of transparency in juridical review was to facilitate meaningful stakeholder participation rather than to be a value in itself.

The indicator loadings of the reasonableness construct differed in size and need further interpretation. If the correlation of a criterion is low, this does not mean that it was not considered relevant for appraisal. Instead, it may not have been a suitable indicator for the degree of reasonableness. Accordingly, in processes where the criteria cost-effectiveness, effectiveness beyond health benefit, and budget impact—indicators with high loadings—are considered, this is likely to reflect reasonable decision making. The reason for the low correlation with the relevance of effectiveness in terms of the health gain (0.31) could be that it had little discriminative power: 84% of respondents stated that effectiveness was relevant. Thus, it appears not to be meaningful to explain the construct's variance. Overall, the AVE of this construct was low (27%). Considering the multitude of health system influences that we neglected, this value can still be considered acceptable.

Limitations

This study has several limitations. First, even if the model suggests causal relationships, these are still of an explorative nature because of the novelty of the research approach. This is consistent with the use of PLS modeling techniques, which are better suited for exploration and extension of existing theories than for theory confirmation. Also, the retrospective study design limits our interpretations. Causality is best established by means of prospective experimental studies, which was not possible in this research context. As in all retrospective observational studies, unmeasured constructs may have influenced the coverage decision processes.

Most studies collected data by extracting decision documentation, which restricts the analysis to well-documented decision processes and may be subject to publication bias. Our approach attempted to collect general indicators of our constructs, which can also be collected for less documented and rigorous decision-making processes.[17] However, using common categories for many processes also limits the level of detail that can be assessed. For example, variation in the level of influence from the stakeholder types was not mapped. We asked for types of interest groups but not the number of institutions involved for each type, which may have conflicting interests. Nevertheless, the reliability and validity measures suggest that the components' variation is captured sufficiently. Although highly correlated, we did not model a link between participation and transparency as the literature does not support a clear causal relationship (i.e., into one direction) between these components. An estimation in which a link going from transparency to participation was inserted showed that the path coefficient between transparency and participation was positive and significant. However, the original SEM estimates did not change in terms of sign and significance.

The estimation is based on a small sample size. We accounted for this by selecting PLS-PM and complying with a rule of thumb for the minimum sample size.[48]

Some decision makers are overrepresented, in particular, decisions from Spain, the UK NICE, or the German FJC. Obtaining responses from countries with less HTA activity and looser regulations is difficult as there are fewer experts in the field.

Nevertheless, the data set can be regarded as representative in terms of technologies and health conditions represented, especially relating to cancer and diseases of the circulatory system.[4]

Most decision makers have focused on coverage of pharmaceuticals that are sufficiently represented. However, our data suggest that also other technologies are increasingly the subject of formal procedures. Regarding nonresponse bias, 5 of 7 decisions that were excluded because of too many missing values dealt with screening, vaccination, or rehabilitation. Given that these may be technologies with less structured processes, they may be underrepresented in our sample.

The distribution of countries that were represented reflects major OCED/EU member countries. Decision making of new EU accession countries is represented to a limited extent because responses only from Poland were obtained.

The sensitivity analyses demonstrated that influences from nonresponse and nesting effects were small. Nevertheless, a multilevel approach would be preferable to correctly identify variation at country, decision maker, and decision levels. If more data were available, unobserved heterogeneity could have been addressed to identify clusters of decision making.[53]

Policy Implications

Transparency, participation, and scientific rigor are procedural criteria for coverage decision processes frequently called for in the literature and required by social law. In case tradeoffs between these exist, this study suggests that, first, rigorous assessment of the evidence and, second, sufficient stakeholder involvement should be promoted to achieve reasonable coverage decisions.

CONCLUSIONS

This is the first study that uses PLS modeling methods to empirically analyze relations between characteristics of decision processes and the reasonableness of substantive criteria used in coverage decisions. Rigorous assessment and intense stakeholder participation appeared effective in promoting reasonable decision making, whereas the influence of transparency was not significant. A sound evidence base seems most important as the degree of scientific rigor of assessment had the strongest effect. Further theoretical and empirical research is necessary to extend the theoretical basis of our hypotheses on the factors influencing decisions and to confirm our results.

ACKNOWLEDGMENTS

We thank Reiner Leidl for valuable comments on the manuscript and development of the survey. We also thank Leon Krefting and Doreen Hoyer for extraction of information on survey respondents. All respondents contributed valuable expertise in the survey on coverage decision processes. We thank those respondents who preferred to remain anonymous, as well as the following who agreed to disclose their names: Lieven Annemans (Belgium), Fernando Antonanzas (Spain), Alun Cameron (Australia), Ch.-Markos Dintsios (Germany), Josep Fuste (Spain), Mendel Grobler (Australia), Iñaki Gutièrrez-Ibarluzea (Spain), Niklas Hedberg (Sweden), Iñaki Imaz (Spain), Ian Jacob (United Kingdom), Jürgen John (Germany), Maria Luisa Lŏpez-Garcia (Spain), Dan Ollendorf (United States), Michelle O'Neill (Ireland), Katherine Payne (United Kingdom), Matthias Perleth (Germany), Maarten Postma (Netherlands), Risto Roine (Finland), Petra Schnell-Inderst (Germany), Pedro Serrano-Aguilar (Spain), Lesley Tilson (Ireland), Dominique Vandijck (Belgium), and Magdalena Wladysiuk (Poland).

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;329(7472):972–5.

2. Sorenson C. Use of comparative effectiveness research in drug coverage and pricing decisions: a six-country comparison. Issue Brief (Commonw Fund). 2010;91:1–14.

3. Henshall C, Mardhani-Bayne L, Fronsdal KB, Klemp M. Interactions between health technology assessment, coverage, and regulatory processes: emerging issues, goals, and opportunities. Int J Technol Assess Health Care. 2011;27(3):253–60.

4. Kanavos P, Nicod E, van den Aardweg S, Pomedli S. The impact of health technology assessments: an international comparison. Euro Observer. 2010;12(4):1–7.

5. Eddy D. Reflections on science, judgment, and value in evidence-based decision making: a conversation with David Eddy by Sean R. Tunis. Health Aff. 2007;26(4):w500–15.

6. Eddy DM. Clinical decision making: from theory to practice: anatomy of a decision. JAMA. 1990;263(3):441–3.

7. Stevens A, Milne R. Health technology assessment in England and Wales. Int J Technol Assess Health Care. 2004;20(1):11–24.

8. Velasco-Garrido M, Kristensen F, Nielsen C, Busse R. Health Technology Assessment and Health Policy-Making in Europe: Current Status, Challenges and Potential. Brussels, Belgium: European Observatory on Health Systems and Policies; 2008.

9. Rogowski W. An economic theory of the fourth hurdle. Health Econ. 2013;22:600–10.

10. Hutton J, McGrath C, Frybourg JM, Tremblay M, Bramley-Harker E, Henshall C. Framework for describing and classifying decision-making systems using technology assessment to determine the reimbursement of health technologies (fourth hurdle systems). Int J Technol Assess Health Care. 2006;22(1):10-18.

11. Persad G, Wertheimer A, Emanuel EJ. Principles for allocation of scarce medical interventions. Lancet. 2009;373(9661):423–31.

12. Cookson R, Dolan P. Principles of justice in health care rationing. J Med Ethics. 2000;26(5):323–9.

13. Daniels N, Sabin J. The ethics of accountability in managed care reform. Health Aff. 1998;17(5):50–64.

14. Friedman A. Beyond accountability for reasonableness. Bioethics. 2008;22(2):101–12.

15. Marckmann G. Health and justice. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2008;51(8):887–94.

16. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999;18(3):341–64.

17. 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.

18. Morgan SG, McMahon M, Mitton C, et al. Centralized drug review processes in Australia, Canada, New Zealand, and the United Kingdom. Health Aff (Millwood). 2006;25(2):337–47.

19. Bryan S, Williams I, McIver S. Seeing the NICE side of cost-effectiveness analysis: a qualitative investigation of the use of CEA in NICE technology appraisals. Health Econ. 2007;16(2):179–93.

20. Fischer KE, Leidl R, Rogowski WH. A structured tool to analyse coverage decisions: development and feasibility test in the field of cancer screening and prevention. Health Policy. 2011;101(3):290–9.

21. Mitton CR, McMahon M, Morgan S, Gibson J. Centralized drug review processes: are they fair? Soc Sci Med. 2006;63(1):200–11.

22. Gauvin FP, Abelson J, Giacomini M, Eyles J, Lavis JN. 'It all depends': conceptualizing public involvement in the context of health technology assessment agencies. Soc Sci Med. 2010;70(10):1518–26.

23. Menon D, Stafinski T. Role of patient and public participation in health technology assessment and coverage decisions. Expert Rev Pharmacoecon Outcomes Res. 2011;11(1):75–89.

24. McMahon M, Morgan S, Mitton C. The Common Drug Review: a NICE start for Canada? Health Policy. 2006;77(3):339–51.

25. Abelson J, Giacomini M, Lehoux P, Gauvin FP. Bringing 'the public' into health technology assessment and coverage policy decisions: from principles to practice. Health Policy. 2007;82(1):37–50.

26. Sorenson C, Drummond MF, Kanavos P. Ensuring Value For Money in Health Care: The Role of Health Technology Assessment in the European Union. Brussels, Belgium: European Observatory on Health Systems and Policies; 2008.

27. Luce BR, Drummond M, Jonsson B, et al. EBM, HTA, and CER: clearing the confusion. Milbank Q. 2010;88(2):256–76.

28. Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. Cost Eff Resour Alloc. 2006;4:14.

29. Sibbald SL, Singer PA, Upshur R, Martin DK. Priority setting: what constitutes success? A conceptual framework for successful priority setting. BMC Health Serv Res. 2009;9:43.

30. Golan O, Hansen P, Kaplan G, Tal O. Health technology prioritization: which criteria for prioritizing new technologies and what are their relative weights? Health Policy. 2011;102(2–3):126–35.

31. Dakin HA, Devlin NJ, Odeyemi IA. 'Yes, 'no' or 'yes, but'? Multinomial modelling of NICE decision-making. Health Policy. 2006;77(3):352–67.

32. 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(5):437–52.

33. Segal L, Dalziel K, Duncan M. Fixing the game: are between-silo differences in funding arrangements handicapping some interventions and giving others a head-start? Health Econ. 2009;19(4):449–65.

34. Erntoft S. Pharmaceutical priority setting and the use of health economic evaluations: a systematic literature review. Value Health. 2011;14(4):587–99.

35. Hair J, Black W, Babin B, Anderson R. Multivariate Data Analysis–A Global Perspective (Global Edition). 7th ed. Upper Saddle River, NJ: Pearson; 2010.

36. Henseler J, Ringle CM, Sinkovics RR. The use of partial least squares path modeling in international marketing. In: Sinkovics RR, Ghauri PN, eds. Advances in International Marketing. Bingley, UK: Emerald; 2009. p 277–320.

37. Rogowski WH, Hartz SC, John JH. Clearing up the hazy road from bench to bedside: a framework for integrating the fourth hurdle into translational medicine. BMC Health Serv Res. 2008;8:194.

38. Stafinski T, Menon D, Philippon DJ, McCabe CJ. Health technology funding decision-making processes around the world: the same, yet different. Pharmacoeconomics. 2011;29(6):475–95.

39. Health Technology Assessment International. HTAi 2010 Abstract Book. 2010 [cited 23 August 2011]. Available from: http://www.htai2010.org/site/

40. International Health Economics Association. 8th World Congress on Health Economics Conference Programme. 2011 [cited 23 August 2011]. Available from: http://www.healtheconomics.org/congress/2011

41. International Health Accounts Team. A System of Health Accounts—Version 2.0, Pre-edited version March 2011. 2011 [cited 27 September 2011]. Available from: http://www.who.int/nha/sha_revision/en/

42. World Health Organization. International Classification of Diseases. 2011 [cited 27 September 2011]. Available from: http://www.who.int/classifications/icd/en/

43. Reinartz WJ, Haenlein M, Henseler J. An empirical comparison of the efficacy of covariance-based and variance-based SEM. International Journal of Research in Marketing. 2009;26(4):332–44.

44. Hair J, Ringle CM, Sarstedt M. PLS-SEM: indeed a silver bullet. J Marketing Theory Pract. 2011;19(2):139–52.

45. Vinzi VE, Trinchera L, Amato S. PLS path modeling—from foundations to recent developments and open issues for model assessment and improvement. In: Vinzi VE, ed. Handbook of Partial Least Square Path Modelling. Heidelberg, Germany: Springer; 2010. p 47–82.

46. Fischer KE. Decision-making in healthcare: a practical application of partial least square path modelling to coverage of newborn screening programmes. BMC Med Inf Dec Making. 2012; 12(1):83.

47. Ringle CM, Wende S, Will A. SmartPLS 2.0. 2005 [cited 28 November 2011]. Available from: http://www.smartpls.com

48. Hair J, Sarstedt M, Ringle CM, Mena J. An assessment of the use of partial least squares structural equation modeling in marketing research. J Acad Marketing Sci. 2012;40:414–33.

49. Tenenhaus M, Vinzi VE, Chatelin Y-M, Lauro C. PLS path modeling. Comput Stat Data Anal. 2005;48(2005):159–205.

50. Culyer AJ. Hic sunt dracones: the future of health technology assessment—one economist's perspective. Med Decis Making. 2012;32(1):E25–32.

51. Mullen PM. Quantifying priorities in healthcare: transparency or illusion? Health Serv Manage Res. 2004;17(1):47–58.

52. Syrett K. Health technology appraisal and the courts: accountability for reasonableness and the judicial model of procedural justice. Health Econ Policy Law. 2011;6(4):469–88.

53. Ringle CM, Sarstedt M, Schlittgen R. Finite mixture and genetic algorithm segmentation in partial least squares path modeling: identification of multiple segments in a complex path model. In: Fink A, Lausen B, Seidel W, Ultsch A, eds. Advances in Data Analysis, Data Handling and Business Intelligence. Berlin: Springer; 2010. p 167–76.