Disagreement on cancer drug decisions in Europe

Short title: HTA Disagreement

Laia Maynou[1,2], John Cairns[3]

[1]London School of Economics and Political Science, Health Policy, United Kingdom

[2] Center for Research in Health and Economics (CRES), Universitat Pompeu Fabra, Spain

[3] London School of Hygiene and Tropical Medicine, United Kingdom

Corresponding author

Dr. Laia Maynou
London School of Economics and Political Science (LSE)
Health Policy
Houghton St, WC2A 2AE, London
e-mail: l.maynou-pujolras@lse.ac.uk

Abstract

Objectives. Despite the efforts of the European Union (EU) to promote voluntary cooperation among Health Technology Assessment (HTA) agencies, different reimbursement decisions for the same drug are made across European countries. The aim of this paper is to compare the agreement of cancer drug reimbursement decisions using inter-rater reliability measures.

Methods. This study is based on primary data on 161 cancer drug reimbursement decisions from nine European countries from 2002 to 2014. To achieve our goal, we use two measures to analyse agreement, in other words, congruency: 1) percentage of agreement and 2) the Kappa Score.

Results. One main conclusion can be drawn from the analysis. There is a weak to medium agreement among cancer drug decisions in the European countries analysed (based on the percentage of agreement and the Kappa score). England and Scotland show the highest consistency between the two measures, showing a medium agreement. These results are in line with previous literature on the congruency of HTA decisions.

Conclusions. This paper contributes to the HTA literature, by highlighting the extent of weak agreement among cancer decisions in Europe.

Key words: drug reimbursement system, Kappa score, inter-rater reliability, congruency.

1. Introduction

Many health care systems have Health Technology Assessment (HTA) processes to inform choices regarding which services and products to pay for from public resources. Drug reimbursement decisions, which usually combine clinical and economic evidence with value judgements, are extremely important, not only to patients but also to manufacturers and health care professionals. Drug reimbursement processes have attracted attention from several authors. In the literature, one can find comparative analyses, i.e. describing a number of different national models [indicatively 1-4], descriptive and comparative studies analysing reimbursement decisions [5-13] and empirical analyses on HTA decisions [14-18].

Despite the efforts of the European Union (EU) to promote voluntary cooperation among Health Technology Assessment (HTA) agencies, reimbursement decisions for the same drug differ across European countries [e.g. 6,8,11,18]. Diverging HTA decisions mean that a particular treatment is covered in one country and not in another. While these decisions can be justified from a policy perspective (well-established procedures), often in neighbouring countries, this may be difficult to comprehend for the public [8], particularly because the clinical evidence reviewed is largely the same, and countries, while not of equal wealth, are of broadly comparable levels of economic development. As a result, one might expect broadly similar decisions (positive or negative) on drug reimbursement to be taken. However, this is not the case [e.g. 6,8,11,18].

To find out why these countries reach different conclusions, in our latest work [18], we made a methodological contribution to determining the factors that might lead to different drug reimbursement decisions on cancer drugs. Results showed significant associations between reimbursement decisions and a drug´s estimated cost-effectiveness, the existence of a financial managed entry agreement (MEA), a health system based on social health insurance, the condition´s incidence rate and the socioeconomic characteristics of the country.

Following on our previous research, our aim in this paper is to compare the agreement of cancer drug reimbursement decisions in a sample of European countries using inter-rater reliability measures. The current study is based on primary data on cancer drug reimbursement decisions from nine European countries from 2002 to 2014: Belgium, Germany, France, Spain, Sweden, Portugal, Poland, England and Scotland. This dataset was collected as part of the ADVANCE-HTA project [11,18]. To achieve our goal, we use two inter-rater reliability measures to analyse agreement or congruency: 1) percentage of agreement and 2) the Kappa Score. The inter-rater reliability is the degree of agreement among

raters. It is a score of how much homogeneity or consensus exists in the decision given by various parties. In this analysis, rater refers to the HTA agencies or systems.

The HTA system's characteristics are the drivers of our findings. Owing to the nature of their system (for a review, see [18]), France, Germany and Spain tend to accept all new drugs into the national system. Other European countries, such as England, Scotland, Poland and Belgium, have high rates of restriction and rejection. Portugal and Sweden have acceptance rates similar to Spain. However, the Spanish system is different from the Portuguese and Swedish because it does not formally allow for rejection and restriction at the national level, but the different regions enjoy some freedom of implementation. The Spanish system introduced the Informe de Posición Terapeutico (IPT), which allows a restricted or rejected decision, at the end of 2013. However, our data pre-date its implementation. In England, until 2016, when a cancer drug reimbursement decision was negative, there was still the option to get it funded through the Cancer Drug Fund (CDF). From 2016 onwards, NICE can recommend entry to the CDF. However, the adoption by the CDF is outside of the HTA process which we are studying.

Previous literature using inter-rater reliability measures to examine agreement among decisions finds, in most cases, low agreement. However, these studies only consider a few countries and/or few drug-indications and/or a short time period [8,12,6,19]. In this paper, we look at nine European countries and we compare decisions for 161 cancer drug-indications from of the period 2002 to 2014. By examining just one therapeutic area, we make the decisions more comparable across the setting. This paper contributes to the HTA literature, shedding light on the weak to medium agreement across cancer drug reimbursement decisions in Europe.

The remainder of the paper is structured as follows. Section 2 describes the methods, including the definition of the database and the applied statistics. The results are presented in section 3, and the main findings are discussed in the final section.

2. Methods

2.1. Database

The current research is based on primary data on cancer drug reimbursement decisions from nine European countries from 2002 to 2014: Belgium, Germany, France, Spain, Sweden, Portugal, Poland, England and Scotland. The database was restricted to these countries to ensure a formal HTA system and public availability of information. The Scottish Medicines Consortium (SMC) was the starting point of our study because it appraises all drugs approved by either the Medicine and Healthcare

Product Regulatory Agency (MHRA) or the European Medicines Agency (EMA). The drugs selected were classified under "malignant disease and immunosuppression" on the SMC website. The list of drugs was validated by reviewing National Institute for Health and Care Excellence (NICE) decisions for any additional observations (especially on further indications for included drugs). This process produced 81 drugs and 161 drug-indications (to account for multiple cancer indications of included drugs, as reimbursement decisions may vary for different indications of the same drug). Thus, 161 cancer drug-indications were considered for each country. Cancer was selected because of the high level of public interest in these reimbursement decisions, and because many cancer drugs have been appraised thus providing a rich dataset. Data was collected during the EU-funded project Advance-HTA (7th Framework Programme Grant No. 305983). The creation of the dataset is explained in detail in two previous studies [11,18]. Supplementary Table 1 presents the data sources.

This database contains the outcome of the decision and the date when the decision was published. The decision outcome describes the final decision regarding the adoption of the technology. We define the outcome as Non-Favourable, Favourable with restrictions and Favourable. These are the three main categories of the HTA decision. To distinguish between "Favourable with restrictions" and "Favourable", the decision is considered to be restricted only when it differs from the indication detailed in the marketing authorisation (e.g. when reimbursement is limited to a sub-population of the patients for whom the drug has been authorized). The rate of acceptance, restriction and rejection for each year in each country are generated from the individual cancer drug decisions collected for our database.

In order to capture all possible decisions, we included another category in the decision variable: Non-assessment. This category can be the result of different circumstances, such as the manufacturer deciding not to apply for reimbursement or the decision-maker not requiring assessment of the drug by the HTA agency.

In order to get robust results, we have to consider drug-reimbursement decisions which are comparable across countries. In other words, the different HTA systems in Europe, incorporate different HTA decisions' definitions. While for Belgium, England, Poland, Portugal, Scotland and Sweden, the outcome of the decision comprises all aspects of the reimbursement and it is used to accept or reject the new technology, for Germany and France, it just indicates the added value of the drug and is used to decide on the type of price agreement. Moreover, for Spain, the national government tends to accept all drugs but then, the regions can decide on the level of access. As a result, in the analysis, we compare across the six comparable countries and then consider the other three (Germany, France and Spain).

2.2. Applied Statistics

We use two inter-rater reliability measures, the percent agreement and the Cohen's kappa score [20]. The percent "agreement among raters" is measured by the number of agreed decisions over the total decisions. It has been widely used when measuring inter-rater reliability as it is easy to calculate and directly interpretable (e.g. [21,6]). However, the percent agreement statistic does not take account of the possibility of chance agreement. So, it may overestimate the true agreement among raters [21]. Cohen [20] suggests the possibility of a false agreement if the agreement is by chance. For instance, this chance agreement can arise if the HTA system tends to accept/restrict/reject all drugs, does not perform an in-depth assessment (e.g. only clinical assessment) or assessments are not comparable. However, McHugh [21] states that if the raters (HTA systems/agencies) are well trained, the percent agreement is a realistic measure to determine interrater reliability.

The Cohen's kappa coefficient, κ, is a statistic which measures inter-rater agreement for qualitative (categorical) items. It ranges from −1 to +1, where 0 represents the amount of agreement that can be expected from random chance, and 1 represents perfect agreement between the raters. While kappa values below 0 are possible, Cohen notes they are unlikely in practice. If the kappa score is negative, it means that the two observers agreed less than would be expected just by chance. As with all correlation statistics, the kappa is a standardized value and thus is interpreted the same across multiple studies [20,21].

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

Where Pr(a) represents the actual observed agreement, and Pr(e) is the proportion of units which would be expected to agree by chance. The denominator, 1 - Pr(e), reflects the maximum agreement beyond chance that would have been possible given the marginal distributions. Note that the sample size consists of the number of observations made across which raters are compared. The statistical significance of the kappa coefficient is based on the chi-square distribution.

Cohen suggests the following interpretation of the kappa coefficient [21]: values ≤ 0 as indicating no agreement and 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement. However, for healthcare research, McHugh [21] argued that the standard interpretation of the kappa score was too mild. For example, 60% agreement implies 40% faulty evidence. For healthcare decisions, 40% of the sample evaluations

disagreed might be considered a serious quality problem. As a result, the author framed a stricter interpretation of kappa values for healthcare: κ=0-0.20 no agreement; κ=0.21-0.39 minimal agreement; κ=0.40-0.59 weak agreement; κ=0.60-0.79 moderate agreement; κ=0.80-0.90 strong agreement; κ>0.90 almost perfect agreement. While McHugh's [21] threshold was defined in a clinical context, it shows that a stricter approach to the evidence is required. This can also hold true for national healthcare decision-making, such as HTA decisions. In the results, both thresholds are used for the interpretation.

However, kappa scores have three limitations or paradoxes [22-24]: 1) for high values of concordance low values of kappa can be recorded, 2) asymmetric, imperfectly imbalanced tables have a higher kappa than perfectly imbalanced and symmetric tables and 3) if the sample of assessed drug-indications is different across countries, it finds values that often lead to underestimation of the actual concordance present in the data. The first two paradoxes will mainly affect countries that have high congruency (in our case those that tend to accept all drugs). Regarding the third paradox, to overcome the limitation, a sample of drug-indications was selected that is common to NICE and SMC, two of the agencies showing a higher agreement [19].

3. Results

Descriptive statistics are presented in Table 1 and Figure 1. This information characterises countries as restrictive or not restrictive, e.g. a country that is not restrictive about giving access (Germany, France, Spain) or, conversely, a country, which tends to restrict and reject access (England, Scotland, Poland and Belgium). The Supplementary Figure 1 reports similar information to Figure 1 but only for the assessed drug-indications (that is, removing the non-assessed category).

Another indicator for the type of HTA system is the non-assessment rate variable. The rate of non-assessment does not change over time because it is the percentage of non-assessed drug-indications out of the 161 drugs for the entire period. In line with the previous variables, Germany, France and Spain have a very low rate, assessing nearly all new drugs. Scotland also joins this group; even though it has a high rate of restriction and rejection, it assesses nearly all drugs. Sweden, Portugal, Poland and England, in contrast, all have a high rate of non-assessment.

The following tables, Table 2 and 3, show the results of the analysis. First of all, we take into account the whole sample (Table 2 and 3), while in the Supplementary Table 2 and 3, we reduce the sample to the drug-indications assessed by NICE and SMC. In these tables, we first present the less restrictive countries (France, Germany and Spain) and then the rest. Results are nearly the same for both samples.

For the interpretation of the results of the percent agreement (Table 2), we follow Allen *et al* [6] thresholds. For the first 6 countries (i.e. Belgium, England, Poland, Scotland and Sweden), results show in most of the cases a low congruence, apart from Belgium-Portugal, Scotland-England, Sweden-Belgium and Portugal-Sweden where there is medium congruence. For France, Germany and Spain, we can see that there is high congruence among them. Portugal shows high agreements with France and Germany, due to the high rate of acceptance in Portugal. However, while for France, the number of assessed drug-indications is high, for Germany, it is low (i.e. due to data availability).

We present the results of the kappa score for the whole sample of 161 cancer drug-indications (Table 3) and for the subsample of 80 drug-indications assessed by NICE and SMC (Supplementary Table 3). To interpret the kappa coefficient the McHugh threshold is stricter than the Cohen one. Table 3 is highlighted in colours based on the McHugh threshold but the legend also reports the Cohen threshold. Results show in most of the cases there is none (McHugh threshold) to slight (Cohen threshold) agreement between countries. The closest agreement that we find, even if it is still considered weak to moderate, is between England and Scotland. There is minimal to fair agreement for Portugal-Belgium, Sweden-Belgium and England-Poland.

4. Discussion

The main aim of this research was to compare the agreement of cancer drug reimbursement decisions in a sample of nine European countries (2002 to 2014) using two inter-rater reliability measures, the percentage of agreement and the Kappa Score.

Our findings present mixed results. While for percentage of agreement the overall results are a medium agreement among cancer drug decisions in the nine European countries analysed, the kappa score shows weak agreement. The percent agreement is easily calculated and directly interpretable. However, it does not take account of the possibility of chance agreement. So, it may overestimate the true agreement among raters [21]. While the kappa score takes it into account, it has some limitations or paradoxes which may lower the estimate of agreement excessively [21]. Overall, taking into account both measures, there is a weak to medium agreement among cancer drug decisions in the European countries analysed.

England and Scotland show the highest consistency between the two measures, showing medium agreement. This result was expected because both countries have a rather similar notion of cost-effectiveness and the incremental cost-effectiveness ratio (ICER) partly (and importantly) determines

their HTA bodies' decision, the cost structure of the health services are similar and the comparators are usually the same. England and Scotland are followed by Portugal-Belgium and Sweden-Belgium but with less agreement. This result might partly be explained by a similar proportion of accepted, restricted and rejected decision among these countries.

For the rest of the countries, there are two main reasons to explain the low agreement. First of all, agreement will depend on the assessing rates of the countries. Some countries assess by law all new drugs (Scotland, Spain, Germany and France) while others have a different system in place (i.e., regarding who initiates the reimbursement process) or a prioritisation policy and, as a result, they have a lower assessing rate. Secondly, the nature of the HTA system matters in the final reimbursement decision, for instance, the system does not allow for restriction and rejection decisions or there is less availability of managed entry agreements [18].

Comparing the percentage of agreement and the kappa score, one can see for some countries (Spain, France and Germany) a high congruence using the percentage of agreement (Table 2) but a non-agreement when using the kappa score (Table 3). The explanation for these findings is that as France and Spain tend to assess all drugs and accept nearly everything, when the kappa is computed with respect to another country that also has a high rate of acceptance, such as Portugal or Sweden, most of the kappa scores imply false or chance agreement, due to the system characteristics and the calculation of the kappa score. Moreover, for Germany, we only had access to few decisions (21) and in all cases, they were positive. For this reason, the kappa score is 0 or impossible to calculate. For these three countries, the kappa score has a limitation as has been shown in previous literature (paradox 1), i.e., for high values of concordance low values of kappa are recorded [22-24]. However, as explained above, the results for France, Spain and Germany must be treated cautiously as they are not directly comparable to the other six HTA systems.

Our findings are in line with previous literature looking at congruency of HTA decisions [8,12,6,19]. Allen *et al* [6] used the percentage of agreement and a taxonomy of systems to assess HTA decisions. However, results demonstrate less alignment between HTA recommendations and HTA Process taxonomic sets. Nicod and Kanavos [8], using the Kappa score, found that 46% of the drug-indication pairs studied (287) received diverging recommendations across countries (five countries) and that the level of agreement between agencies was poor to moderate. Nicod [12] while looking only at 10 orphan drugs, found that six out of ten drugs received diverging HTA recommendations. Kanavos *et al.* [19] also used the Kappa Score and showed a better level of agreement in assessing health technologies in a pharmaceutical sample (20 drugs) than in a medical devices and other technologies sample. So, while

most of the previous studies found low to moderate agreement among HTA decisions, results differ depending on the drug sample size, the number of countries and the therapeutic area. In our analysis, we focused on one therapeutic area (i.e. to make it comparable), on 161 cancer drug-indication pairs, nine countries (taking into account two main sets) and a longer period of analysis 2002-2014. This paper contributes to the HTA literature, by highlighting the extent of weak to medium agreement among cancer decisions in Europe.

During the recent years, there has been much effort at the European level to promote voluntary cooperation among the HTA agencies. In 2004, EUnetHTA was created and there is currently a new proposal for regulation of HTA being discussed in the EU Parliament and the Council [25]. The proposed regulation aims to promote the convergence of methods, tools, assessments and cooperation among HTA agencies based on a joint clinical assessment. The main objective is to reduce duplication and improve the efficiency of the HTA systems. However, even if the EU is trying to have binding decisions at the European level, the countries/regions are still concerned about their own competencies. So, despite substantial efforts at the EU level to promote voluntary cooperation and sharing of assessment, HTA decisions for cancer drugs differ across the EU countries analysed. These diverging decisions can be justified from a policy perspective (as they are based on well-established procedures, from a patient or societal perspective), however differential access to a particular treatment across neighbouring countries may be difficult to understand for the public, particularly in countries with comparable levels of health care spending and similar levels of income [8]. However, further research is needed to determine the effect of this differential access on health outcomes.

Although the results are satisfactory, during this research we encountered a number of challenges. One limitation is related to data collection and availability. For instance, for Germany, we only had access to few decisions. Moreover, to combine the data from each country in a single analysis some assumptions were needed (i.e. details on the assumption can be found in [18]). Regarding the methods, while the Kappa score is a standard inter-rater reliability measure, it has several limitations. According to McHugh [21], the Cohen threshold for the kappa score in healthcare is too mild. We overcome this limitation by using both thresholds in the results interpretation.

In this study, we have only looked at cancer drugs. Future research should also compare across therapeutic areas, in order to explore the level of agreement among decisions.

5. References

[1] **Franken M, Le Polain M, Koopmanschap M, Cleemput I.** (2012) Similarities and differences between five European drug reimbursement systems. *Int J Technol Assess Health Care* **28**(4), 349–357.

[2] **Barneih L, Manns B, Harris A, et al.** (2014) A synthesis of Drug Reimbursement Decision-Making Processes in Organisation for Economic Co-operation and Development countries. *Value in Health* **17**, 98-108.

[3] **Franken M, Stolk E, Scharringhausen T, de Boer A, Koopmanschap M.** (2015) A comparative study of the role of disease severity in drug reimbursement decision making in four European countries. *Health Policy* **119**(2),195–202.

[4] **Oortwijn W, Determann D, Schiffers K, et al.** (2017) Towards Integrated Health Technology Assessment for Improving Decision Making in Selected Countries. *Value in Health*, **20**(8),1121-1130.

[5] **Fischer K.** (2012) A systematic review of coverage decision-making on health technologies-Evidence from the real world. *Health Policy* **107**, 218-230.

[6] **Allen N, Liberti L, Walker SR, Salek S** (2007) A Comparison of Reimbursement Recommendations by European HTA Agencies: Is There Opportunity for Further Alignment?. *Front Pharmacol*, **308.** doi: 10.3389/fphar.2017.00384

[7] **Clement F, Harris A, Li J, Yong K, Lee K, Manns B** (2009) Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada. *JAMA* **302**(13), 1437–1443.

[8] **Nicod E, Kanavos P** (2012). Commonalities and differences in HTA outcomes: A comparative analysis of five countries and implications for coverage decisions. *Health Policy* **108**,167–177.

[9] **Franken MG, Nilsson F, Sandmann FG, Boer AD, Koopmanschap MA** (2013) Unravelling Drug Reimbursement Outcomes: A Comparative Study of the Role of Pharmacoeconomic Evidence in Dutch and Swedish Reimbursement Decision Making. *PharmacoEconomics* **31**,781-797.

[10] **Grepstad M, Kanavos P** (2015) A comparative analysis of coverage decisions for outpatient pharmaceuticals: evidence from Denmark, Norway and Sweden. *Health Policy* **119**(2),203–11.

[11] **Maynou L, Cairns J** (2015) Why do some countries approve a cancer drug and others don't? Journal of Cancer Policy **4**, 21-25.

[12] **Nicod E** (2017) Why do health technology assessment coverage recommendations for the same drugs differ across settings? Applying a mixed methods framework to systematically compare orphan drug decisions in four European countries. *Eur J Health Econ* **18**(6),715-730

[13] **Allen NJ, Walker SR, Liberti LE, Salek S** (2017) Health Technology Assessment (HTA) Case Studies: Factors Influencing Divergent HTA Reimbursement Recommendations in Australia, Canada, England, and Scotland. *Value in Health* **20**(3),320-328.

[14] **Dakin H, Devlin N, Odeyemi I** (2006) ''Yes'', ''No'' or ''Yes, but''? Multinomial modelling of NICE decision-making. *Health Policy* **77**, 352–367.

[15] **Cerri K, Knapp M, Fernandez JL** (2014) Decision making by NICE: examining the influences of evidence, process and context. *Health Econ Policy Law* **9**, 119-41.

[16] **Dakin H, Devlin N, Feng Y, et al.** (2015) The Influence of Cost-effectiveness and Other Factors on NICE Decisions. *Health Econ* **24**,1256-71.

[17] **Charokopou M, Majer IM, Raad J, et al.** (2015) Which Factors Enhance Positive Drug Reimbursement Recommendation in Scotland? A Retrospective Analysis 2006–2013. *Value in Health* **18**(2), 284-91.

[18] **Maynou L, Cairns J** (2019) What is driving HTA decision-making? Evidence from cancer drug reimbursement decisions from 6 European countries. *Health Policy* **123**(2):130-139.

[19] **Kanavos P, Fontrier AM, Habl C, et al.** (2017) Study on impact analysis of Policy Options for strengthened EU cooperation on Health Technology Assessment (HTA). EU Publications, 2017. [Accessed 14 February 2018: https://ec.europa.eu/health/sites/health/files/technology_assessment/docs/2018_ia_policyoptions_en.pdf]

[20] **Cohen J** (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas*. **20**,37–46.

[21] **McHugh M** (2012) Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* **22**(3): 276–282.

[22] **Feinstein AR, Cicchetti DV** (1990) High agreement but low kappa: I. The Problems of Two Paradoxes. *J Clin Epidemiol* **43**,543-548.

[23] **Tang W, Hu J, Zhang H, Wu P, He H** (2015) Kappa coefficient: a popular measure of rater agreement. *Shanghai Arch Psychiatry* **27**(1), 62–67.

[24] **Flight L, Julious SA** (2015) The disagreeable behaviour of the kappa statistic *Pharm Stat* **14**,74-78.

[25] **Proposal for a Regulation (EC) No 2018/0018(COD)of the European Parliament and of the Council on health technology assessment and amending Directive 2011/24/EU** (2018). [Accessed 14 February 2018 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018PC0051#document1]

Acknowledgements

This paper was developed with data collected during the research project funded by the European Commission's Research Framework Programme (FP7), Advance-HTA (no305,983). The authors would like to thank the Advance-HTA Consortium members and the experts from the National HTA Agencies for their support on the data collection process. Moreover, the authors appreciate the good feedback received in the different conferences where this piece of research was presented (AES and Spanish HTA meeting).

Table 1. Descriptive statistics: variables of interest (2002-2014)

Country	Rate of acceptance	Rate of restriction	Rate of rejection	Rate of non-assessment
Belgium	65.79 (24.06)[1]	32.10 (21.06)	2.11 (4.21)	19.25
	(12.5, 100)[2]	(0, 75)	(0, 12.5)	
France	90.93 (12.00)	2.88 (4.76)	6.19 (11.07)	4.35
	(60, 100)	(0, 12.5)	(0, 40)	
Germany	100 (0)	0 (0)	0 (0)	0
	(100, 100)	(0, 0)	(0, 0)	
Portugal	86.28 (17.27)	5.13 (11.04)	8.59 (12.49)	72.67
	(50, 100)	(0, 33.33)	(0, 33.33)	
Poland	60.27 (43.52)	25.76 (31.50)	13.97 (15.49)	41.61
	(0, 100)	(0, 87.5)	(0, 44.44)	
Spain	87.74 (23.35)	12.26 (23.35)	0 (0)	9.94
	(16.67, 100)	(0, 83.33)	(0, 0)	
Sweden	87.13 (17.52)	6.76 (11.14)	6.11 (9.39)	68.32
	(55.56, 100)	(0, 33.33)	(0, 22.22)	
Scotland	25.00 (18.69)	47.84 (31.08)	27.17 (21.02)	7.2
	(0, 50)	(7.14, 100)	(0, 54.55)	
England	46.27 (30.46)	25.91 (20.52)	27.82 (28.41)	36.36
	(0, 100)	(0, 66.67)	(0, 87.5)	

[1]mean (standard deviation), [2] (max,min). Source: ADVANCE-HTA dataset, created by the author

Figure 1. Acceptance, restriction and rejection rate + non-assessed rate (2002-2014)

Source: ADVANCE-HTA dataset, created by the author

Table 2. Percentage of agreement for 161 cancer drug-indications (2002-2014)

Agreement in HTA decisions (assessed drugs)	Spain	Germany	France	Sweden	Scotland	Portugal	Poland	England	Belgium
France	86.18%[1] (123)	94.74% (19)		72% (50)	28.83% (111)	79.55% (44)	24.18% (91)	39.33% (89)	59.84% (122)
Germany	57.14% (14)			68.75% (16)	46.67% (15)	100% (5)	0% (13)	50% (12)	43.75% (16)
Spain				72.09% (43)	22.83% (92)	73.17% (41)	26.83% (82)	31.98% (74)	58.25% (103)
Belgium				65.12% (43)	40.40% (99)	65.85% (41)	34.18% (79)	39.51% (81)	
England				39.39% (33)	63.75% (80)	44.83% (29)	45.16% (62)		
Poland				18.92% (37)	37.84% (74)	31.58% (38)			
Portugal				71.43% (21)	38.24% (34)				
Scotland				30.23% (43)					
Sweden									

Notes: [1] *percent of agreed decisions and in brackets, the total number of common assessed decisions in both countries.*

>=75%	High Congruence	>=50% & <75%	Medium Congruence	<50%	Low Congruence

Table 3. Kappa Score for 161 cancer drug-indications (2002-2014)

KAPPA SCORE	Spain	Germany	France	Sweden	Scotland	Portugal	Poland	England	Belgium
France	0.1441**	0.0000		-0.0219	0.0165	0.1520**	0.0188	-0.0050	0.0712**
Germany	0.0000			0.0000	0.0000	N/A	0.0000	0.0000	0.0000
Spain				0.1254	-0.0483	0.0322	0.0328	-0.0492	0.0229
Belgium				0.2483**	0.1348***	0.2807***	0.0400	0.0834	
England				-0.0015	0.4661***	0.1111	0.2075***		
Poland				-0.0054	0.0130	0.1534***			
Portugal				-0.1053	0.0165				
Scotland				-0.0644					
Sweden									

*Notes: significance levels: *** p<0.01, ** p<0.05, * p<0.1*

McHugh Threshold	*Cohen Threshold*	*Kappa coefficient*
No Agreement	**No Agreement**	**≤ 0.0**
	None to Slight	**0.01-0.20**
Minimal Agreement	**Fair Agreement**	**0.21 - 0.39**
Weak Agreement	**Moderate Agreement**	**0.4 - 0.59**
Moderate Agreement	**Substantial Agreement**	**0.6 - 0.79**
Strong Agreement	**Almost Perfect Agreement**	**0.8 – 0.9**
Almost perfect Agreement	**Almost Perfect Agreement**	**> 0.9**