

This is a repository copy of *Understanding Variations in Relative Effectiveness:A Health Production Approach*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/94890/>

Version: Accepted Version

Article:

Towse, Adrian, Jonsson, Bengt, McGrath, Clare et al. (5 more authors) (2016)
Understanding Variations in Relative Effectiveness:A Health Production Approach.
International Journal of Technology Assessment in Health Care. pp. 363-370. ISSN 0266-4623

<https://doi.org/10.1017/S0266462315000719>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Understanding variations in relative effectiveness: a health production approach

Journal:	<i>International Journal of Technology Assessment in Health Care</i>
Manuscript ID	IJTAHC-14-025.R3
Manuscript Type:	Methods
Date Submitted by the Author:	n/a
Complete List of Authors:	Towse, Adrian; Office of Health Economics, Jönsson, Bengt; Stockholm School of Economics, Economics McGrath, Clare; AstraZeneca, Payer & Real World Evidence Mason, Anne; University of York, Centre for Health Economics Puig-Piero, Ruth; Servei Català de la Salut (CatSalut),, Mestre-Ferrandiz, Jorge; Office of Health Economics, Pistollato, Michele; Office of Health Economics, Devlin, Nancy ; Office for Health Economics
Keywords:	Relative Effectiveness, Drug Evaluation, European Union
Abstract:	<p>Abstract</p> <p>Background: Relative effectiveness has become a key concern of health policy. In Europe this is because of the need for early information to guide reimbursement and funding decisions about new medical technologies. However, ways that effectiveness (does it work?) and efficacy (can it work?) might differ across health systems are poorly understood.</p> <p>Methods: This paper proposes an analytical framework, drawing on production function theory, to systematically identify and quantify the determinants of relative effectiveness and sources of variation between populations and health care systems. We consider how methods such as stochastic frontier analysis and data envelopment analysis using a Malmquist productivity index could in principle be used to generate evidence on, and improve understanding about, the sources of variation in relative effectiveness between countries and over time.</p> <p>Results: Better evidence on factors driving relative effectiveness could: inform decisions on how to best use a new technology to maximum effectiveness; establish the need if any for follow up post-launch studies, and provide evidence of the impact of new health technologies on outcomes in different health care systems.</p> <p>Conclusions: The health production function approach for assessment of relative effectiveness is complementary to traditional experimental and observational studies, focusing on identifying, collecting and analysing data at the national level, enabling comparisons to take place. There is a strong case for exploring the use of this approach to better understand the impact of new medicines and devices for improvements in health outcomes.</p>



SCHOLARONE™
Manuscripts

For Peer Review

Understanding variations in relative effectiveness: a health production approach**Short title: *a health production approach to relative effectiveness***

Adrian Towse¹, Bengt Jonsson², Clare McGrath³, Anne Mason⁴, Ruth Puig-Peiro⁵, Jorge Mestre-Ferrandiz¹, Michele Pistollato¹ and Nancy Devlin¹

¹Office of Health Economics, ²Stockholm School of Economics, ³AstraZeneca, ⁴University of York, ⁵Servei Català de la Salut (CatSalut)

Abstract: 238; Main Paper 4370

Key words: Relative Effectiveness; Drug Evaluation; European Union.

Abstract

Background: Relative effectiveness has become a key concern of health policy. In Europe this is because of the need for early information to guide reimbursement and funding decisions about new medical technologies. However, ways that effectiveness (does it work?) and efficacy (can it work?) might differ across health systems are poorly understood.

Methods: This paper proposes an analytical framework, drawing on production function theory, to systematically identify and quantify the determinants of relative effectiveness and sources of variation between populations and health care systems. We consider how methods such as stochastic frontier analysis and data envelopment analysis using a Malmquist productivity index could in principle be used to generate evidence on, and improve understanding about, the sources of variation in relative effectiveness between countries and over time.

Results: Better evidence on factors driving relative effectiveness could: inform decisions on how to best use a new technology to maximum effectiveness; establish the need if any for follow up post-launch studies, and provide evidence of the impact of new health technologies on outcomes in different health care systems.

Conclusions: The health production function approach for assessment of relative effectiveness is complementary to traditional experimental and observational studies, focusing on identifying, collecting and analysing data at the national level, enabling comparisons to take place. There is a strong case for exploring the use of this approach to better understand the impact of new medicines and devices for improvements in health outcomes.

Acknowledgements: Research funding was received from Pfizer. We would like to thank three anonymous referees for their constructive comments. We are also grateful to Rikard Althin for comments on the Malmquist Index. All remaining errors and omissions are the responsibility of the authors.

BACKGROUND

Relative - or comparative - effectiveness has become a key health policy concern. Payers are increasingly interested in assessing the relative effectiveness of new health care technologies relative to standard care or other designated comparators (1). This reflects growing recognition of the need to understand how efficacy demonstrated in a randomised controlled trial (RCT) will translate into added benefit in routine clinical use in different health care settings. Relative effectiveness evidence is also important when assessing whether results in one jurisdiction can be applied elsewhere. Regulators want to use it in post-launch benefit risk assessments (2). In the case of the anti-obesity drug rimonabant, for example, such an assessment suggested both that relative effectiveness was poorer than relative efficacy and safety risks were higher than expected. The product was subsequently withdrawn by the European Medicines Agency (EMA) (3).

Relative effectiveness can be defined as “the extent to which an intervention does more good than harm compared to one or more alternative interventions under the usual circumstances of health care practice” (4). This contrasts with *relative efficacy*, which is a comparison “under ideal circumstances”, i.e. “under clinical trial conditions” (5). Comparative effectiveness is very similar to relative effectiveness, and is defined as “comparing health outcomes and the clinical effectiveness, risks, and benefits of two or more medical treatments, services and items” (6) which for Garber and Sox means “real world settings” (7).

We focus on the European Union (EU) to explore the translation of relative efficacy into relative effectiveness in recognition of current efforts to develop EU-level approaches to assessment.

There appears to be a general view that relative efficacy of most therapies will be the same across Europe in most circumstances - although very few studies have tested that assumption (8). In the RE-LY trial, the relative efficacy of a new oral anticoagulant varied between countries even under RCT conditions, depending on the efficiency of warfarin management (9). Thus even relative efficacy can differ between countries when health care practice varies.

How relative effectiveness might differ across health systems within the EU is poorly understood. Yet variations in relative effectiveness have implications for:

1. The extent to which national and regional pricing and reimbursement (P&R) decisions vary, and why. This will depend on whether decisions about patient access to medicines depend on considerations of relative effectiveness and, if so, the extent to which relative efficacy is accepted as a proxy for it;
2. The desire to assess benefit-risk by the EMA beyond launch into the “product life cycle” in which expected benefits are compared with expected risks, based on evidence from actual use (10);
3. The work of the European network for Health Technology Assessment (EUnetHTA) in particular Work Package 5 (WP5), to standardise methods and data requirements for relative effectiveness assessments across the Member States, including the applicability of evidence from one setting to another (11), following the recommendation of the High Level Pharmaceutical Forum in 2008;
4. The EU drive to increase cross border health care in Europe: the implication is that citizens should obtain same health outcomes regardless of the country where they obtain a treatment;

5. The actions health systems can take to be sustainable and meet health needs in the face of economic downturn (12). In the US, the purposes of conducting CER include providing both the basis to improve care quality and enable savings (13).

Health technology industries face a paradigm shift if they are to be paid and regulated on the basis of 'does it work?' (effectiveness) instead of 'can it work?' (efficacy) (14). Industry remuneration and regulatory benefit-risk assessment would also depend on the actual capacity, competence, or other ability of the health care system to provide access to and otherwise appropriately deliver the drug, i.e., the effectiveness of the system, as well as the potential effectiveness of the drug. This has two consequences. Firstly, there is an increased demand for studies to identify local impact. In the extreme, if a company has to plan to conduct different relative effectiveness studies for 28 EU Member States post-launch the implications for development costs may be substantial, even if done in collaboration with health care systems. Secondly, evidence of relative effectiveness in routine use will need to be integrated into a range of decisions: not only decisions on reimbursement or regulatory status, but also decisions about any required changes to current care practices needed if the benefits of new technologies are to be optimised.

METHODS

This paper focuses on how we can identify whether and to what extent relative effectiveness differs across EU health care systems, and what data and methods are needed to study the impact of changes over time when new treatment options are introduced. We undertook a brief scoping review to identify approaches to measure relative effectiveness. We focused on literature reviews and empirical studies that distinguished between efficacy and effectiveness, i.e. outcomes in clinical practice. We then explored the potential relevance of an analytical framework drawing on production function theory to consider how certain sets of inputs and processes yield specified outcomes. This approach can, in principle, provide a systematic way of identifying and quantifying the determinants of relative effectiveness, and sources of variation between populations and health care systems. We set out the elements of a multilevel model in this paper. In an accompanying article (15), we use this model to categorise evidence on breast cancer to illustrate potential differences in the relative effectiveness of a new drug entering the market.

The production function approach is complementary to traditional experimental and observational studies, and focuses on identifying, collecting and analysing data at the national level for assessment of relative effectiveness. Our approach consisted of four steps. First, we set out an analytical framework. The health production function approach uses health as the output of interest. This reflects the EU Directive on cross border care which forms the basis for the coordination of Health Technology Assessment (HTA) and relative effectiveness assessment in Europe (16). Relative effectiveness is the basis for assessing therapeutic added value and cost-effectiveness at the national level. What determines value is the **absolute** difference in effectiveness, which translates for example into the number of QALYs gained. Second, we classify inputs (or 'factors') into three main levels: patient, provider, and the health care environment or system. The relative effectiveness of a drug is the additional net output (health) achieved by adding a new drug to usual care or substituting it for another treatment. Third, the health production function provides an approach to organize the data and an analytical framework to examine the differences in relative effectiveness across jurisdictions. We briefly outline the types of model that might be appropriate

and consider its role in relation to the increasing demand for observational studies of the impact of the use of drugs in practice. Lastly, we consider the policy implications of our analysis.

RESULTS

Understanding the reasons for differences in relative effectiveness

Eichler et al. (1) set out the methodological options and challenges in measuring relative efficacy. Their subsequent comprehensive review (5) identified two reasons for an “efficacy-effectiveness gap” : biological (including genetics, age, sex, co-morbidity and baseline severity of disease) and behavioural (e.g. variation in adherence to the treatment regimen). Economists would add that any understanding of a potential “efficacy-effectiveness gap” requires translating trial outcomes into long term health outcomes. However other factors will also play a role, notably the health system. For example, the health system and the environment impact on disease severity and co-morbidity. Adherence is not only behavioural, but also depends on the organisation and resources used for managing adherence, and thus the efficiency of the healthcare system. The selection of patients for treatment and the management of patients over time will depend on incentives, organisation and resources in the health care system. These elements may differ across jurisdictions and will impact on long term outcomes.

Estimating relative effectiveness: empirical methods

Currently relative effectiveness, as opposed to relative efficacy, is treated in the literature and assessed by HTA and reimbursement agencies in one of three ways:

- (a) it is assumed to be the same as relative efficacy, and any potential “efficacy-effectiveness gap” is ignored, i.e. modelling is based on clinical trial data;
- (b) it is estimated using decision analytical modelling. However, this technique is typically used to project longer term health outcomes from short term or intermediate efficacy measures, or to make indirect comparisons, rather than to model the impact of the health system on the translation of the efficacy of a technology into outcomes in routine care. An intermediate case is when the relative risk reduction from a trial is applied to the absolute risk level for different groups of patients in different countries (17). However this approach does not take into account other factors that may influence relative effectiveness, for example differences in adherence between countries;
- (c) observational studies or pragmatic clinical trials are proposed as methods to explore outcomes in routine clinical practice within a health care system (17-19). The number of observational studies being undertaken within health systems is increasing, but they can be time consuming and expensive unless relevant data are already routinely captured by a health care system. Analysis can be complex and results open to differing interpretations, although methods to tackle confounding and other issues are improving. Pragmatic trials also can be time consuming and expensive unless infrastructure investment is put in place and recruitment streamlined.

None of these approaches are routinely used to detect and explain differences in relative effectiveness across health systems. A review of the literature (8) found no observational studies investigating variations in relative effectiveness across Member States, although there were safety

studies and studies comparing other aspects of care across several Member States. It is therefore not possible to address the question of differences in relative effectiveness on the basis of current evidence.

One route forwards is for more observational studies or pragmatic clinical trials to be conducted across Member States. Given the cost and complexity of conducting them, however, it would make sense to explore another approach which can use existing data, and could be used to help identify whether differences in relative effectiveness were likely to be of an order to merit further study.

The conceptual framework: A health production function model of the determinants of relative effectiveness

The relative effectiveness of a new technology is likely to depend on how it is combined with other health inputs, and a range of other variables that influence the overall relationship between inputs and health outcomes in the underlying 'health production function'.

Assuming the comparator treatment is one of the inputs in the initial health production function, then relative effectiveness is the change that results when the new treatment is introduced relative to that obtained from the existing (or usual) treatment (and other inputs). In this context 'usual care' in the health system will be a key determinant of the relative effectiveness of any new treatment. The underlying research questions are: if a medicine for, say, breast cancer is granted Market Authorisation by the EMA, indicating that the new technology is available in all EU countries, do we expect that its relative effectiveness will be different across them? If so, why?

One way to investigate influences on relative effectiveness is to treat this analytically as a multilevel model, separating the different levels of influence. Influences may operate on one or more of three main levels: patient, provider and the health care environment and system (see Table 1):

- 'Individual' patient level factors refer to the characteristics of the patients being treated, including their demographic, genetic and clinical characteristics. Not only may the patient mix gaining access to the technology differ but underlying population health may also differ between health systems. The same technology in the same setting may therefore produce different health effects depending on patient characteristics. Even if *relative* efficacy is the same across two health systems, if baseline risk differs then size of the *absolute* incremental effect measured by relative effectiveness may differ.
- The characteristics of the providers. The same technology may produce different health effects on patients with similar characteristics depending on how it is used and combined with other technologies in clinical practice (e.g. screening policy, or treatment setting; see Table 1). Clinical practice may vary in other ways that influence treatment outcomes (both from new technologies and comparators) for example due to economic and other incentives for providers.
- Health care environment and system factors. The availability or otherwise of technologies within the health care system and how technologies are used (and hence the health effects on patients) (i.e. 'comparators' in Table 1) is often driven by P&R policies and clinical guidelines from regional or national decision makers. Waiting times for referral and then access to specialist care also vary between (and within) countries.

TABLE 1 HERE

Some determinants of outcome operate and interact at several levels. Five year survival rate for breast cancer may differ across countries due (in part) to differences in all-cause mortality rates. These may be due to individual-level effects (e.g. smoking status), to environmental effects (e.g. taxes on cigarettes), or to health system public health measures restricting smoking. The key is to identify factors driving outcomes and ensure they are included, but only once. In this example, the variable for smoking would be most appropriately classified and included in the analysis as an individual level effect. In the accompanying article (15), we provide an overview of factors affecting breast cancer outcomes identified from a literature review, grouping them according to the multilevel approach: 'individual level', 'provider level', and 'environment and health care system level' set out in Table 1. The results are set out in Table 2 of Puig-Peiro et al. (15)

The Malmquist DEA model for measuring relative effectiveness

The theory and methods for estimating health production functions have developed rapidly over the last two decades. Important contributions include the development of econometric approaches such as the stochastic frontier approach (20; 21), and also data envelopment analysis (DEA) (22; 23) including the use of Malmquist indexes which allow use of multiple input, multiple output models without prices. The latter is particularly important in a European health care perspective, where prices varies between countries, and often are not possible to observe and calculate in public health care systems. In addition, for some inputs to the health production function, such as individuals time and efforts used for prevention and treatment prices are not directly observable, as is also the case for the output health. Using index numbers that include prices for measurement of changes in total factor productivity, i.e. relative effectiveness is not an option.

A further advantage of the Malmquist approach is that, once the production technology is estimated, either as a production frontier or its dual the cost frontier, one can decompose the change in two component parts: efficiency change and technical change. The methodology proposed by Färe et al (1994)(24) uses DEA to calculate distance functions to produce the Malmquist total factor productivity index, and then decomposes this into technical and efficiency components respectively. This decomposition is important since the change in relative effectiveness in a given countries is made up by a combination of changes in the frontier, and how the country is moving in relation to the frontier.

Figure 1 illustrates how a Malmquist DEA model could help identify and explain potential cross-country variations in the relative effectiveness of the introduction of a new treatment for breast cancer treatment. It can be used to study and decompose the productivity changes both in terms of technical change (best practice change, shown by a shift in the frontier) and technical efficiency change (countries moving closer to the frontier), i.e. to understand how the introduction of a new technology changes outcomes, and hence to establish its relative effectiveness. This is essential for the design of policies aimed at improving relative effectiveness.

For illustrative purposes we assume just two inputs, but in practice the estimation of these models can include multiple inputs and patient attributes such as those set out in Table 1 that will impact on the patient outcomes. The horizontal axis measures the number of "physician visits" as a proxy for the intensity of monitoring and non-drug treatment, and the vertical axis measures the number of

doses of “drugs” as a proxy for the intensity of drug treatment. We could also have used for example the share of visits to a specialist on the horizontal axis and an index of “vintage”, i.e. year of introduction, for drugs. For illustration we show the various combinations of the two variables that yield an 85% five year survival rate. The Production Frontier in Figure 1 is data-driven: it is the envelope of the most efficient data points observed from all countries in the group. We compare the average performance of two of the countries, A and B.

FIGURE 1 HERE

In period t , before the introduction of the new drug, efficiency is not the same in the two countries: e_A is closer to the production frontier in period t than is e_B . They use different combinations of input factors, which may reflect differences in relative prices between drugs and physician visits in the two countries. The distance function Oc/Oe represents the technical inefficiency in country A and B respectively in period t , i.e. how much the use of drugs and visits could have been reduced if they had applied the best practice use implied by the production frontier at period t .

When a new drug or procedure is introduced in period $t + 1$, the production frontier shifts inwards, indicating an improvement in technical efficiency to produce the same level of health outcome with fewer of each of the inputs. The frontier may not change by the same proportion for both countries implying that one country may increase its efficiency in a larger proportion than the other after the introduction of the new drug.

The distance functions Oa/Of shows the new degree of inefficiency, i.e. actual performance compared with best practice on the radial from the origin for the two countries respectively

For each observation, the change from period t to $t+1$ in efficiency can be computed as the ratio between the two distance functions, and the technical change as the geometric mean of the product of two ratios. These are defined in the key to figure 1. The total productivity change, or change in relative effectiveness is the product of the two ratios. All observations can be aggregated to an overall productivity change, using the geometric mean.

The optimal mix of drugs and physician visits that would occur once the new drug was available along the production frontier in period $t+1$ depends on their relative prices and can be studied separately.

We show, in Online Supplementary file 1, the potential use of this approach with a very simple calculation and illustration of the technical efficiency with which combinations of breast cancer screening and treatment with trastuzumab are used in seven European countries.

Strengths and weaknesses of the approach

Using such an approach to empirically identify and explain differences in relative effectiveness may encounter some challenges in practice. For example, it would require data on all relevant factors to be available in a manner that facilitates cross-country comparisons. Further, depending on the specific factors identified as relevant to understanding relative effectiveness in a specific disease area, relationships between explanatory variables (e.g. adherence may depend on socioeconomic status) can confound identification of their relationship to outcomes. However, such challenges are common to many studies and do not preclude useful and valuable research being undertaken.

Valid and reliable price data for (non-drug) inputs and outputs are generally not available, so techniques to estimate cross- and time-series variations in the relationship between inputs and outputs without price data, such as the approach illustrated in Figure 1, are an important step forward. They have been extensively used to understand the relative efficiency of providers (usually hospitals) within a single health care system and to compare the performance of health care systems – see Hollingsworth and Wildman (25), Greene (26) and the EUROHOPE project (27). Jacobs et al. (28) and Coelli et al. (29) provide a comprehensive review of the main econometric approaches for modelling health production functions.

Questions for the decision makers, based on such an analysis, should include (i) whether a country can get closer to best practice and so improve relative effectiveness? (ii) whether the likelihood of gaps between relative efficacy and relative effectiveness mean that post-launch studies are required, and (iii) whether potential differences between countries A and B mean that separate post-launch studies are needed in each?

We are not proposing econometric and economic modelling as an *alternative* to a prospective observational study or pragmatic trial to understand differences across health care systems in the relative effectiveness of a technology. Rather we see this analytical tool as complementary: one that can use existing data (including data already available from post-launch studies) to explore the extent to which the performance of the same technology differs, or might be expected to differ, across settings as a result of differences in populations and health care systems.

Policy implications

There are prima facie reasons why we might expect some convergence in relative effectiveness across the EU over time: the diffusion of medical technology is becoming global and there is a flow of information about best practice. It is also possible to observe some convergence of health care systems in terms of financing, organisation, and performance. Yet our findings suggest it is reasonable at the moment to expect meaningful and persistent variations in relative effectiveness between and within countries. Although the literature indicates that current understanding of what might be important is limited, each case is likely to be driven by a few key factors. The health production function is a tool for identifying these at an early stage and making predictions of their likely quantitative impact. But it can also serve as a basis for targeting observational studies and pragmatic clinical trials post launch, to test predictions and gain further insight into how to best make use of the new technology, including the potential advantages and disadvantages of using coverage with evidence development and risk sharing arrangements to assess and reward relative performance.

There are four broad policy implications from our analysis.

First, at a time when we are seeing that countries increasingly look at P&R decisions and HTA assessments in other countries, the debate around the transferability of relative effectiveness evidence for P&R decisions could be better informed by an understanding of what factors in a health care system are likely in practice to impact on the incremental performance of a technology.

Second, the increasing demand for post launch studies to address uncertainty about relative effectiveness of drugs needs to be put in the context of such an understanding. Studies are

expensive and complex and unnecessary duplication of effort could waste time and resources and delay patient access to effective treatments. A collaborative approach between industry and governments (and in the EU across the 28 Member States) is required to:

- a) Limit the number of extra studies. An improved understanding of the drivers of relative effectiveness can help prioritise these studies. They can be limited to those needed to explore key expected differences between health systems that could impact patient outcomes.
- b) Collect and provide access to necessary data on the relevant contextual factors. At present, access to routine data (for example from registers) is often restricted. In Sweden, a public inquiry described the registers as “a gold mine” for collecting evidence to improve health care and health. Such registers are, however, often oriented to answering specific clinical questions and need to be augmented with a wider set of relevant factors and outcomes for the analysis of relative effectiveness.

Third, where differences in relative effectiveness reflect variation in the efficiency of health care systems, good practice can be identified and shared. Decision makers in poorly performing health systems, who are responsible for the procurement of new technologies and for managing health care providers, can learn from good practice elsewhere. Relative effectiveness is the outcome of an interaction between a technology, patients and the health system (including the providers). The role of the latter and how to improve efficiency is often missing from debates such as that about developing an EU assessment of relative effectiveness. Failure to improve health system performance risks wasted resources or poorer health outcomes.

Fourth, if industry remuneration and regulatory benefit-risk assessment increasingly depend on the actual effectiveness of the health care system in using a drug, as opposed to relative efficacy at launch, then we may see an increasing use of ‘pay for performance’ and risk sharing agreements. Such agreements are, in effect, seeking to estimate the health production function and link payment or use in some way to the outcome. Without an understanding of the contextual factors there is a danger of companies taking the “risk” for factors that are really the responsibility of the health system and vice versa. The potential for using post-launch follow up studies to reduce the time and cost of drug development before market authorization in some areas (cancer being an example where fast access is important and risks can be managed) needs to be informed by an understanding of the key factors driving relative effectiveness.

CONCLUSIONS

We conclude that the step from relative efficacy to relative effectiveness can be substantial and is likely to vary across health systems. Relative efficacy is a function of the attributes of the technology and target patient population. Relative effectiveness additionally depends on the performance of the health care delivery system and wider environmental factors. The potential for differences in relative effectiveness across EU Member State health systems gives rise to a number of public policy issues. The health production function approach is complementary to traditional experimental and observational studies, focusing on identifying, collecting and analysing data at the national level for assessment of relative effectiveness, enabling comparisons to take place. There is a strong case for exploring the use of this approach to enable the differences to be estimated and addressed and any requirements for post-launch observational or experimental studies to be targeted.

For Peer Review

References

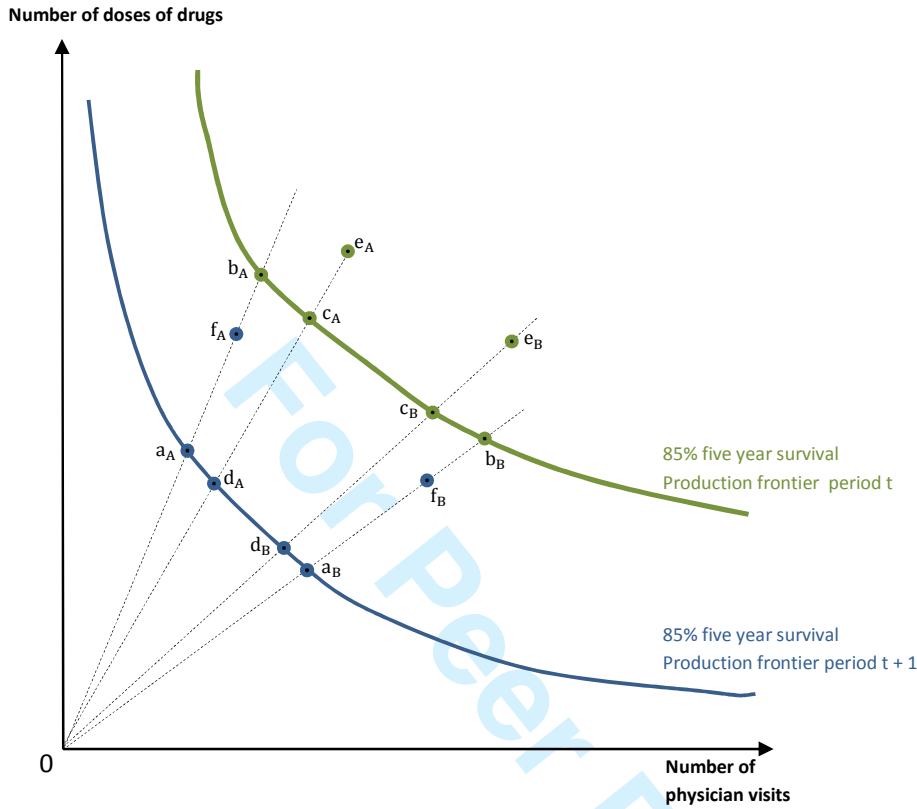
1. Eichler H.-G., Bloechl-Daum B., Abadie E., et al. Relative efficacy of drugs: an emerging issue between regulatory agencies and third-party payers. *Nat Rev Drug Discov*, 2010, 9(4), 277-91.
2. Eichler H.-G., Pignatti F., Flamion B., Leufkens H. & Breckenridge A. Balancing early market access to new drugs with the need for benefit/risk data: a mounting dilemma. *Nat Rev Drug Discov*, 2008, 7(10), 818-26.
3. European Medicines Agency. Assessment Report For Acomplia. EMEA/65105/2009. London; 2009. p. 5.
4. High Level Pharmaceutical Forum. Core principles on relative effectiveness. Brussels: European Commission: Healthcare Industries Working Group on Relative Effectiveness; 2008. p. 10.
5. Eichler H.-G., Abadie E., Breckenridge A., et al. Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. *Nat Rev Drug Discov*, 2011, 10(7), 495-506.
6. The Patient Protection and Affordable Care Act. PL. No. 111-148 sec. 6301; 2010.
7. Garber A.M. & Sox H.C. The role of costs in comparative effectiveness research. *Health Affairs*, 2010, 29(10), 1805-11.
8. Mestre-Ferrandiz J., Puig-Peiró R. & Towse A. Researching the relative efficacy and relative effectiveness of medicines across Europe. OHE Consulting Report. London: Office of Health Economics; 2010.
9. Wallentin L., Yusuf S., Ezekowitz M.D., et al. Efficacy and safety of dabigatran compared with warfarin at different levels of international normalised ratio control for stroke prevention in atrial fibrillation: an analysis of the RE-LY trial. *Lancet*, 2010, 376(9745), 975-83.
10. European Medicines Agency. Road map to 2015: The European Medicines Agency's contribution to science, medicines and health. London; 2011. p. 28.
11. European network for Health Technology Assessment. EUnetHTA Guideline. Levels of Evidence: applicability of evidence in the context of a relative effectiveness assessment of pharmaceuticals. Final version (February). Copenhagen, 2013.
12. Saltman R.B., Calltorp J. & de Roo A.A. Health Sector Innovation and Partnership. In: OECD, (ed.) *Health Reform: Meeting the Challenge of Ageing and Multiple Morbidities*. Paris and Washington, D.C: OECD Publishing, 2011, 143-72.
13. Garber A.M. US Healthcare Reform: Monumental Health System Transformation or Fatally Flawed Compromise? OHE Annual Lecture. London: OHE, 2010.
14. Luce B.R., Drummond M., Jonsson B., et al. EBM, HTA, and CER: clearing the confusion. *Milbank Quarterly*, 2010, 88(2), 256-76.
15. Puig-Peiro R., Mestre-Ferrandiz J., Mason A., et al. Understanding variations in relative effectiveness: applying a health production approach to breast cancer. *Int J Technol Assess Health Care*; (submitted).
16. Jönsson B. Relative effectiveness and the European pharmaceutical market. *Eur J Cancer*, 2011, 12(2), 97-102.
17. Lindgren P., Jönsson B. & Yusuf S. Cost-effectiveness of clopidogrel in acute coronary syndromes in Sweden: a long-term model based on the CURE trial. *Journal of Internal Medicine*, 2004, 255(5), 562-70.
18. Lauer M.S. & D'Agostino R.B., Sr. The randomized registry trial--the next disruptive technology in clinical research? *N Engl J Med*, 2013, 369(17), 1579-81.
19. Luce B.R., Drummond M.F., Dubois R.W., et al. Principles for planning and conducting comparative effectiveness research. *J Comp Eff Res*, 2012, 1(5), 431-40.
20. Zuckerman S., Hadley J. & Iezzoni L. Measuring hospital efficiency with frontier cost functions. *Journal of Health Economics*, 1994, 13(3), 255-80; discussion 335-40.
21. Worthington A.C. Frontier efficiency measurement in health care: a review of empirical techniques and selected applications. *Medical Care Research & Review*, 2004, 61(2), 135-70.

22. Hollingsworth B., Dawson P.J. & Maniadakis N. Efficiency measurement of health care: a review of non-parametric methods and applications. *Health Care Management Science*, 1999, 2(3), 161-72.
23. Hollingsworth B. Non-parametric and parametric applications measuring efficiency in health care. *Health Care Management Science*, 2003, 6(4), 203-18.
24. Färe R., Grosskopf S., Lindgren B. & Roos P. Productivity Developments in Swedish Hospitals: A Malmquist Output Index Approach. In: A. Charnes, W.W. Cooper, A.Y. Lewin & L.M. Seiford, (eds.) *Data Envelopment Analysis: Theory, Methodology, and Applications*. New York: Springer Science, 1994, 253-72.
25. Hollingsworth B. & Wildman J. The Efficiency of Health Production: Re-estimating the WHO Panel Data Using Parametric and Non-parametric Approaches to Provide Additional Information. *Health Econ*, 2003, 12(6), 493-504.
26. Greene W. Distinguishing between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organization's Panel Data on National Health Care Systems. *Health Econ*, 2004, 13(10), 959-80.
27. European Commission. EuroHOPE European Health Care Outcomes, Performance and Efficiency. <http://cordis.europa.eu/projects/241721>; 2010.
28. Jacobs R., Smith P.C. & Street A. *Measuring Efficiency in Health Care: Analytic Techniques and Health Policy*. Cambridge and New York: Cambridge University Press, 2006.
29. Coelli T.J., Rao D.S.P. & O'Donnell C.J. *An Introduction to Efficiency and Productivity Analysis*. Second ed. New York: Springer Science + Business Media, 2005.

Table 1 Overview of factors that may influence the relative effectiveness of a medicine or other technology

Influence level	Category	Variables
Individual / patient level factors	Demographic characteristics	age; gender; socioeconomic status; education; insurance status; employment status; lifestyle factors (e.g. smoking status)
	Clinical characteristics	disease severity, including disease stage at diagnosis; co-morbidity; procedures received; general health and life expectancy; genetic type
	Other characteristics	compliance /concordance; health literacy/awareness
Provider level factors	Provider characteristics	specialisation; private/public status; teaching status; clinical practice (e.g. hospital policy on use of off-label drugs); capital / labour availability (e.g. physician skills, CT scanner); patient selection.
Environment/health care system level factors	Population health	lifestyles, mortality rates, life expectancy, disease prevalence; valuation of outcomes (e.g. EQ-5D values)
	National /regional guidelines / regulations	clinical guidelines; national service frameworks/national plans; legal framework (e.g. funding for some drugs /services may be mandatory)
	Service delivery and organisation	screening, treatment settings, provision of palliative care
	Access issues (local regional/national)	pricing and reimbursement policies ; insurance and co-payment policies ; waiting time targets; comparator drug(s) available
	Economy	GDP; % GDP spent on healthcare; % private funding for healthcare

Figure 1. Illustration of the Malmquist Index applied to the identification of possible variations in the relative effectiveness of breast cancer treatments.



Key		
	Country A	Country B
Performance before the introduction of the new drug (period t)	e_A	e_B
Performance after the introduction of the new drug (period t + 1)	f_A	f_B
Degree of inefficiency before the introduction of the new drug	$\frac{0c_A}{0e_A}$	$\frac{0c_B}{0e_B}$
Degree of inefficiency after the introduction of the new drug	$\frac{0a_A}{0f_A}$	$\frac{0a_B}{0f_B}$
Efficiency change	$\frac{0a_A}{0f_A} \cdot \frac{0e_A}{0c_A}$	$\frac{0a_B}{0f_B} \cdot \frac{0e_B}{0c_B}$
Technical change*	$\left(\frac{0b_A}{0a_A} \cdot \frac{0c_A}{0d_A}\right)^{1/2}$	$\left(\frac{0b_B}{0a_B} \cdot \frac{0c_B}{0d_B}\right)^{1/2}$
Change in productivity (relative effectiveness)*	$\left(\frac{0a_A}{0f_A} \cdot \frac{0e_A}{0c_A}\right) \cdot \left(\frac{0b_A}{0a_A} \cdot \frac{0c_A}{0d_A}\right)^{1/2}$	$\left(\frac{0a_B}{0f_B} \cdot \frac{0e_B}{0c_B}\right) \cdot \left(\frac{0b_B}{0a_B} \cdot \frac{0c_B}{0d_B}\right)^{1/2}$
* There is one Malmquist productivity index defined for each observation. The overall change can be defined as the geometric mean of all observations		

Online Supplementary File 1

Illustrative simple example of the use of a health production frontier

We show the potential use of the health production frontier approach with some easy to follow illustrative calculations of the technical efficiency with which seven European countries use combinations of two inputs, breast cancer screening and trastuzumab treatments, to achieve the output of breast cancer survival.

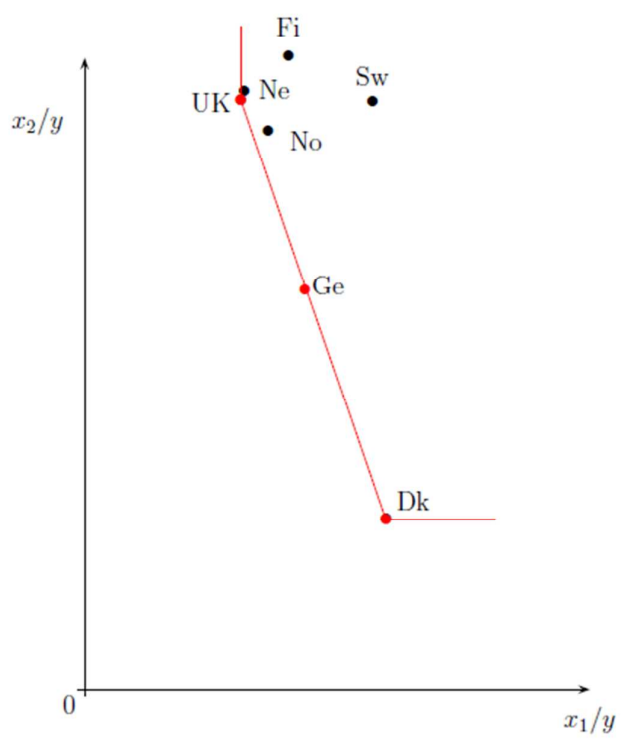
The example is highly simplified because we assume that the output is driven only by combinations of two inputs. We also assume constant returns to scale —an increase of all the inputs at once in a given proportion increases the output exactly in the same proportion— and strong disposability of inputs —it is possible to produce at least the same amount of output with more inputs. This could mean that not all additional inputs necessarily increase outputs (for example, if trastuzumab is given to a woman who is not HER2 positive it will not increase her survival prospects).

Table 1 below sets out for each country and Figure 1 illustrates the position of each country in relation to the frontier. The frontier is defined by the countries achieving an input efficiency index of 1. These are the Denmark, Germany and the UK. We can see that, in this simple illustration, Sweden, for example, appears to be technically inefficient. It uses similar amounts of screening per unit of survival as the UK, but more trastuzumab per unit of survival.

Table 1. Farrell Input Based Efficiency Scores Under Constant Returns to Scale and Strong Disposability of Inputs

Country	$F_i(y, x)$	x_1 Trastuzumab	x_2 Screening	y_1 Survival
Sweden	0.80	1950	80	86.0
Denmark	1.00	1900	22	81.5
Norway	0.97	1500	75	84.7
Finland	0.90	1600	86	85.7
UK	1.00	1300	74	79.2
Netherlands	0.99	1400	80	84.5
Germany	1.00	1600	53	83.6

Figure 1: Technical Efficiency



Review