

Consulting Report

Researching the Relative Efficacy and Relative Effectiveness of Medicines Across Europe

Analysis and Implications for the Future of HTA in Europe

Jorge Mestre-Ferrandiz, OHE
Ruth Puig-Peiró, OHE
Adrian Towse, OHE

25 October 2010

This consulting project was funded by the following companies:

Amgen
Astra Zeneca
Bayer Healthcare
Lilly
Merck
Novartis
Novo Nordisk
Johnson & Johnson
Merck
Pfizer
Roche
Sanofi Aventis

We are grateful for their financial support and also for comments received on an earlier draft of the report. We would particularly like to thank Clare McGrath of Pfizer who chaired the Steering Group of the participating companies.

This report was written by the following OHE Consulting team members:

Dr Jorge Mestre-Ferrandiz
Ruth Puig-Peiró
Professor Adrian Towse

with advice, comment and support from:

Professor Bengt Jönsson, Stockholm School of Economics
Anne Mason, Centre for Health Economics, University of York
Professor Nancy Devlin, OHE Consulting

Table of Contents

EXECUTIVE SUMMARY	1
1 INTRODUCTION AND CONTEXT.....	9
1.1 DEFINITIONS.....	10
1.2 OBJECTIVES AND DELIVERABLES OF THE OHE CONSULTING STUDY.....	11
1.3 STRUCTURE OF THE REPORT	13
2 CURRENT USE OF EFFECTIVENESS DATA IN SELECTED EU COUNTRIES.....	14
2.1 LINKAGES BETWEEN KEY VARIABLES: FRAMEWORK	14
2.2 LINKAGES BETWEEN KEY VARIABLES: COUNTRY BY COUNTRY ANALYSIS.....	15
2.3 PRICING AND REIMBURSEMENT / MARKET ACCESS: GENERAL OVERVIEW.....	23
2.4 METHODS OF RELATIVE EFFECTIVE / EFFECTIVENESS ASSESSMENT	28
2.5 PROCESS OF RELATIVE EFFICACY / EFFECTIVENESS ASSESSMENT	34
2.6 IMPLICATIONS FOR A PAN-EUROPEAN ASSESSMENT (OF RELATIVE EFFICACY).....	37
3 AN OVERVIEW OF PAN-EUROPEAN INITIATIVES RELATED TO HTA.....	38
3.1 THE POLICY CONTEXT IN EUROPE	38
3.2 THE EUROPEAN NETWORK FOR HEALTH TECHNOLOGY ASSESSMENT (EUNETHTA).....	39
3.2.1 <i>Results from the EUnetHTA Project 2006-2008.....</i>	<i>39</i>
3.2.2 <i>The HTA Core Model.....</i>	<i>40</i>
3.2.3 <i>The EUnetHTA Joint Action 2010-2012.....</i>	<i>42</i>
3.3 THE SWEDISH EU PRESIDENCY INITIATIVE ON RELATIVE EFFECTIVENESS	44
3.4 NEXT STEPS	47
3.5 REFERENCES	47
4 SYSTEMATIC LITERATURE REVIEW ON RELATIVE EFFECTIVENESS IN EUROPE	49
4.1 INTRODUCTION	49
4.2 METHODS.....	52
4.3 RESULTS	53
4.4 FINDINGS AND IMPLICATIONS.....	66
4.5 REFERENCES	68
5 PUTTING THE JIGSAW TOGETHER WITH AN INDUSTRY VISION OF THE FUTURE	70
5.1 INTRODUCTION	70
5.2 SUMMARISING OUR FINDINGS	70
5.3 WHAT COULD BE THE BENEFITS OF PAN-EUROPEAN ASSESSMENT?	71
5.4 RELATIVE EFFECTIVENESS AND THE POST-LAUNCH ENVIRONMENT	71
5.5 THE DANGERS OF OVERLY HIGH STANDARDS	72
APPENDIX 1. CURRENT USE OF RELATIVE EFFECTIVENESS IN SELECTED EU MEMBER STATES	74
APPENDIX 2. SEARCH STRATEGIES FOR THE LITERATURE REVIEW.....	109
APPENDIX 3. KEY POINTS FROM EICHLER, ET AL. (2010)	111

Executive Summary

Objectives of the Study

Ten member companies of the EFPIA HTA Task Force commissioned OHE Consulting to provide research, data and analysis that would enable EFPIA both to support and develop its policy position and to contribute positively to the emerging dialogue around the development of HTA within Europe. This study also will enable the pharmaceutical industry to work through how some form of a pan-European approach to the assessment of relative efficacy / relative effectiveness might work in practice in a way that could be acceptable to both Member States and the pharmaceutical industry.

Our work was designed to:

1. Describe and critically review current approaches to the assessment of relative efficacy/relative effectiveness in Member States
2. Identify and measure the underlying sources of variation of relative efficacy / effectiveness across Europe by reviewing the literature currently available on this question
3. Look at the feasibility and desirability of a pan-EU relative efficacy / effectiveness assessment in terms of its generalisability across Member States and its impact on patient access to medicines
4. Consider how EFPIA could develop arguments in response to these findings to advance its position on pan-European HTA and engage constructive dialogue around the scientific merit and public policy desirability of pan-European approaches
5. Identify additional research that would enable the industry to both develop a stronger evidence base to support its position and be seen to contribute positively to the dialogue around the practical issues and desirability of any harmonisation

Definitions

Throughout the report, we have used the following definitions for the key variables:

Efficacy: The extent to which an intervention does more good than harm under ideal circumstances. (This comparison is usually based on data from clinical trials, and can be made at launch).

Relative efficacy: The extent to which an intervention does more good than harm compared to one or more alternative interventions under ideal circumstances. (This comparison is usually based on data from clinical trials, and can be made at launch).

Effectiveness: The extent to which an intervention does more good than harm under the usual circumstances of health care practice. (This comparison needs to be based on data collected in clinical practice over time).

Relative Effectiveness: 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. (This comparison needs to be based on data collected in clinical practice over time).

Cost effectiveness: Cost-effectiveness analysis (CEA) is a form of economic analysis that compares the relative costs and outcomes (effects) of two or more courses of action for a defined patient group and indication.

Generalisability refers to the extent whether the results of an HTA report can be readily adopted by other settings.

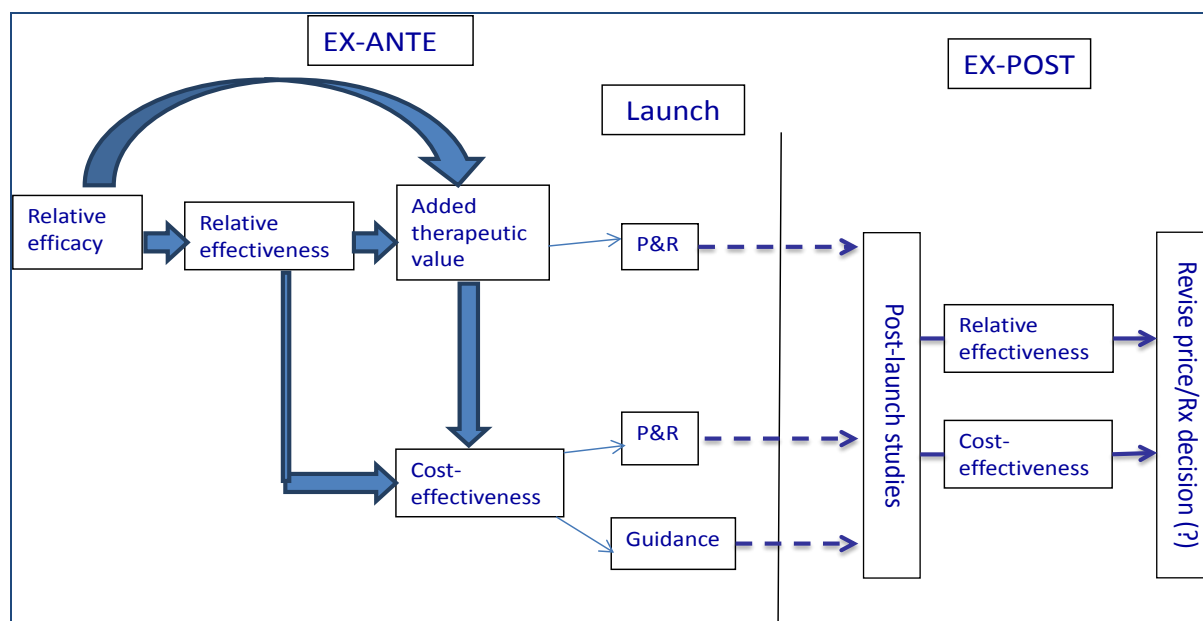
Transferability refers to the ability to apply information and/or data from one report for the user's target setting. The more context specific the data are, the less transferable they are.

Note, however, that there is no agreed definition of 'added therapeutic value' in the HPLF Report or any subsequent literature and its use is different in different countries.

Current use of relative efficacy and effectiveness data by selected Member States

The objective of this part of the work was to synthesise how (if at all) EU countries use (relative) efficacy/effectiveness data within their health technology assessment (HTA) and/or pricing and reimbursement (P&R) mechanisms. The focus is on eight countries: England and Wales, France, Germany, Italy, the Netherlands, Scotland, Spain and Sweden. This list includes the 'big 5' markets plus three countries where HTA is important.

Figure ES1. Framework to explore linkages between key variables include post-launch studies



The key consideration is to understand the relationship between (relative) efficacy, relative effectiveness, measures of added therapeutic value/degree of innovativeness and cost effectiveness within HTA and P&R mechanisms, both at launch (i.e. ex-ante) and after launch (ex-post). Figure ES1 outlines the possible inter-relations between the different variables, and shows all the possible transformations that can occur within any one country. It also shows the possibility of manufacturers needing to undertake post-launch studies.

Several transformations can occur prior to launch (under 'EX-ANTE'):

- Efficacy to effectiveness
- Efficacy to added therapeutic value
- Effectiveness to added therapeutic value
- Effectiveness to cost effectiveness

Based on these transformations (however done), the relevant body will then make a final decision regarding coverage (reimbursement), pricing and/or recommendations on use ('guidance') (under 'LAUNCH').

Figure ES1 also includes the possibility of the use of post-launch studies in decision making, based on collecting evidence on either relative efficacy/effectiveness and/or cost effectiveness. As a result of the new evidence, the relevant authorities might revise the price and/or the reimbursement status of drugs, or alternatively, keep their price/reimbursement status as it is.

Several models were identified for the use of relative efficacy data across the (selected) Member States:

- Relative efficacy => Added therapeutic value => P&R (France, Germany)
- Relative efficacy => Relative effectiveness => Added therapeutic value => P&R (Italy, Spain)
- Relative efficacy => Relative effectiveness => Cost-effectiveness
 - => P&R (Sweden, the Netherlands)
 - => 'Guidance' (England and Wales, Scotland)

Based on this desk research, it seems to us that:

- Whilst there appears to be commonality around the use of relative efficacy, based on RCTs, as a starting point, it is unclear whether the methods used to identify, include, and analyse RCTs are similar.
- It is also unclear to what extent the comparators used in any assessment of relative efficacy might differ, and the methods used to identify, include and analyse trials including indirect comparators.
- In other words it is not clear if the relative efficacy assessment is the same (or very similar) across MSs.
- The use made of a relative efficacy assessment seems, however, to be quite different across MSs. Even if the relative efficacy assessment is the same, how it is currently used to (ultimately) determine patient access to medicine is not the same.

Initiatives for a pan-European approach

There are a number of initiatives at European level, including the High Level Pharmaceutical Forum process which initiated the reflection on pan-European approaches to relative efficacy/effectiveness, the 2009 Swedish Presidency Initiative for cross border collection of observational data, the EULAR example of such a cross-border initiative in the field of drugs for RA, the MEDEV group of Member State P&R bodies, and the EUnetHTA programme. Of immediate interest is the work planned for 2010-2012 under EUnetHTA's Joint Action Programme. It is essential that the industry is engaged in the various pan-European initiatives that are taking place. However, there appears to be a lack of clarity as to exactly which issues EUnetHTA expects to be raised within the Joint Action Programme and exactly where such an exercise might get to. In addition the EMA continues, as set out in its Road Map to 2015, to seek to expand on: the EPAR, setting out for MS HTA/P&R bodies its analysis of RCT data submitted to it; its role on offering scientific advice alongside MS HTA/P&R bodies; and its plans for post launch management of benefit/risk (which in theory involves looking at post launch benefit as well as risk). In the field of orphan drugs, the Commission has issued a tender for work to create a pan-European exchange of knowledge on new orphan drugs to facilitate early access at the MS level.

Systematic literature review findings on underlying efficacy /effectiveness differences across MSs

The primary objective of this literature review was to find and understand the extent of likely variation if any in relative efficacy and relative effectiveness of drugs used in one or more of the 27 Member States.

Potential Sources of Variation

The potential factors leading to differences in reported treatment effects can be grouped as follows:

1. Patients' characteristics (e.g. demography, case mix, co-morbidity) or disease features (e.g. severity, epidemiology) may be different. These two elements are linked. The point is that the patients are heterogeneous. If the groups being compared are different in some key attribute then the reported response to treatment may be different.
2. Studies may use different clinical endpoints or health outcomes measures. The studies may be measuring different things. Additionally the valuation of a given health state may differ between countries – thus the outcome is the same but its value to the patient (or to a sample of the population) is different.
3. The comparator or related features of the comparator intervention may be different.
4. Others sources of clinical practice may vary in a way which might influence treatment outcomes. The literature includes many factors under this term including differences in:
 - Prescription (differences in DDD) and related aspects of drug utilisation

- Concordance /compliance with treatment
- The number and type of tests, length of stay in hospital, number of outpatient visits, and the location of treatment e.g. primary versus secondary care.

Findings regarding relative efficacy

The literature testing for heterogeneity in relative efficacy and relative effectiveness across European countries is very scarce. We found ten clinical trial based studies, one of which was a simple RCT and nine which were cost-effectiveness studies using RCT evidence on clinical data. Our main findings from these studies were:

- The degree of heterogeneity remains an unresolved empirical issue.
- Clinical practice variation is one of the most cited factors when analysing heterogeneity of cost effectiveness results but only from the cost side. It is not clear that differences in clinical practice necessarily impact on relative efficacy or relative effectiveness.

The reason for so few studies (or perhaps the consequence) is that there is a common assumption in the literature that:

- Relative efficacy is constant and therefore generalisable across settings. When measured as relative risk reduction, it does not vary with the baseline risk.
- As a consequence most studies are underpowered to test for heterogeneity and therefore it is not detected.

Regression techniques as Multilevel Models using RCTs data are starting to be used to explore between country variations in both costs and treatment effects. Such studies as have been published suggest that relative efficacy does not vary, and indeed the published applications of the techniques reduce such (non-significant) differences as are found.

Findings regarding relative effectiveness

Our findings were that:

- Between country heterogeneity on relative effectiveness is more likely to occur than on relative efficacy because many of the potential sources of variation (notably around patient heterogeneity) are more likely to be found in observational than RCT studies.
- No observational studies met our search criteria for exploring relative effectiveness across different Member States. The absence of reported studies directly exploring this issue is a matter of concern. More efforts are needed to produce cross-border observational studies collecting and analysing registry data.
- One review, highlighted at the 2009 Swedish conference, used safety rather than effectiveness data from three EULAR registers (UK, Sweden and Germany) to compare the relative risk of infections of TNF.

- Some evidence from cardiovascular diseases shows that prescribing behaviour (affected by clinical guidelines and reimbursement conditions) and adherence to treatment were factors that may drive variations.

Putting the jigsaw together – what are the options for EFPIA?

Benefits of a pan-European assessment of relative efficacy

On relative *efficacy*, a pan-European approach may be useful and feasible as most HTA/P&R bodies use relative efficacy. A number of issues need to be taken into account:

- Whilst there appears to be commonality around the use of relative efficacy, based on RCTs, as a starting point, it is unclear whether the methods used to identify, include, and analyse RCTs are similar. Particular issues are the selection of end points for consideration and the treatment of indirect comparisons, where head-to-head RCTs do not exist.
- Two benefits could include avoiding duplication of effort and raising quality of assessment in some MSs. However, ultimately the benefits will depend on the methods to assess RCT evidence, especially on the use of indirect comparators. There is a danger that unrealistically high standards will be set in terms of either trial end points or the need for direct head-to-head comparisons.
- It should be noted also that the use made of a relative efficacy assessment seems, however, to be quite different across MSs. Even if the relative efficacy assessment is the same, how it is currently used to (ultimately) determine patient access to medicine is not the same. It is therefore quite possible that a pan-European assessment of relative efficacy may make no difference to the time taken to review a new drug by a MS or to patient access. Indeed if standards are set too high, then requirements for longer pre-launch RCTs may delay access.

Relative effectiveness post-launch

On relative *effectiveness*, there are important differences across the MSs HTA /P&R bodies when translating relative efficacy into relative effectiveness. In addition, more empirical evidence measuring heterogeneity on relative effectiveness is needed to understand the extent of transferability of data.

However, there is a danger for the industry in allowing the current assumption to continue that relative effectiveness varies across MSs because of differences in clinical practice. Our thinking is that most differences in clinical practice may impact on resource use and therefore on cost, but are less likely to impact on relative effectiveness. Such differences in relative effectiveness as do exist are more likely to be the result of differences in the characteristics of the patients treated, and in the comparator used. Both of these points are linked to the issue of the uptake of new technologies. Historic uptake rates influence comparators and new uptake rates influence the mix of patient groups that are treated in a particular MS.

This question becomes more important if we assume that the industry's model for the EU (and US) will be to seek to launch products earlier and shift data requirements to post-launch environment. Companies may, of course, try to put more effort in pre-launch to avoid future uncertainty. The EMA is also indicating, however, that it is likely to seek more data post launch to support benefit/risk assessments. A key requirement will be to be able to generalise any effectiveness results post-launch. If this is not the case, there is a risk of requirements being imposed for multi-country post-launch data requirements. In the extreme companies would need to conduct post-launch studies in all 27 MS, and in some cases different studies may be required for safety monitoring and for looking at effectiveness. This is simply not commercially or practically feasible in the absence of a transformation in the use of electronic patient records.

Industry co-operation with other stakeholders

In addition to the necessity of involvement in the key EUnetHTA work programmes including the continuation of WP4 on the HTA Core Model and the new WP5 on the Relative Effectiveness Assessment of Pharmaceuticals, the industry needs to engage with a number of the key stakeholders including the:

- Key MSs, both Health Ministries and HTA/P&R bodies
- MEDEV (Medicine Evaluation Committee for Pharmaceutical Policy Cooperation between the Statutory Health Insurance Institutions in Europe), another key grouping of MS bodies
- Swedish Presidency initiative on Relative Effectiveness
- EMA

To this end EFPIA could hold a Workshop or a series of Workshops designed to bring key stakeholders together to discuss important aspects of pan-European assessment of relative efficacy/effectiveness. This could provide a platform to discuss progress on different initiatives such as that of the Swedish Presidency and the EUnetHTA programme, and/or to debate particular methodological and practical issues such as the use of indirect comparisons, and the reasons why relative effectiveness might vary between MSs. Such Workshops could also explore how such pan-European initiatives might improve patient access to innovative medicines. To support such a programme of Workshops, or indeed to support EFPIA's policy development, some additional research could be commissioned.

Options for future research to support EFPIA

On the current use of relative efficacy and effectiveness data

We have looked at how in principle information on relative efficacy and relative effectiveness is used in key MSs. A next logical stage would be to explore particular case studies for a sample of drugs to examine how assessments are done in practice in different MSs and the implications for access to medicines across Europe. Are drugs deemed to offer additional therapeutic value in one country and not in another (i.e. before considerations of price and cost-effectiveness and considered). The

objective would be to understand differences in methods. It would help in particular to include both views on the suitability of particular end-points, and the use of indirect comparisons. These are likely to be two of the key issues for the pharmaceutical industry in any common approach to relative efficacy assessment. It will also help to identify any differences in comparators across the different MSs.

We have indicated that it would make sense to supplement the analysis of written material with interviews of key staff from HTA bodies to understand the thinking behind particular decisions.

Understanding more about relative effectiveness differences across Europe

We have argued that there is need for the industry to better understand factors that in reality drive differences in relative effectiveness across MSs. As we noted, that most differences in clinical practice may impact on resource use and therefore on cost, but are less likely to impact on relative effectiveness. Such differences in relative effectiveness as do exist are more likely to be the result of differences in the characteristics of the patients treated, and in the comparator used. Both of these points are linked to the issue of the uptake of new technologies. Historic uptake rates influence comparators and new uptake rates influence the mix of patient groups treated in a particular MS.

A study designed to explore this could look at:

- A literature review, perhaps focused on one or more of the following disease areas (some cancers, cardiovascular disease, diabetes, osteoporosis, or Venous thromboembolism), to identify papers not picked up by our general review
- Some empirical research, using vignettes describing indication /treatment combinations and clinical panels in several countries to explore the sources of any differences in treatment patterns. This could be linked to existing EFPIA commissioned work looking at underlying variations in the uptake of medicines.

1 Introduction and Context

EFPIA is developing its position on HTA within Europe and is discussing its position on whether or not to support moves for a pan-European assessment of relative efficacy and/or relative effectiveness. Our understanding of its position following the EFPIA Board Workshop on HTA, 21 June 2010, is as follows:

1. Accepting that national payers and HTA bodies wish to assess the added value of new medicines and that sound and holistic HTA with proper involvement of patients, physicians and industry has the potential to stimulate innovation and patient access to new medicines.
2. Recognising that EU-level actions on the assessment of relative efficacy and effectiveness could add value if they tackle unnecessary duplication among the 27 Member States and enable greater consistency and clarity, lead to a raising of standards of methodological and process aspects in HTA, improve predictability, and contribute to better and more timely access of medicines to patients.
3. EFPIA could support a EU mechanism for the (clinical) assessment of relative efficacy (at launch) and relative effectiveness (over time) subject to a number of core understandings and principles that would need to be in place:
 - Regulatory approvals - based on quality, safety and efficacy – must remain separate from any assessment of relative efficacy (at launch) or relative effectiveness (over time). EFPIA accepts the current mandate of EMA (as laid down in the legislation), including improved EPARs and better articulation of how the Agency makes its benefit-risk assessments. EMA already has the power to mandate comparative trials, but beyond this EFPIA will resist any expansion of the role of EMA in the area of relative efficacy and relative effectiveness.
 - All cost considerations and economic evaluations should remain at the national level.
 - Such assessments should be recognised and accepted in subsequent national decisions on pricing, reimbursement and access.
4. It is too early to say what body or what mechanism would be best suited for such EU assessments.
5. Currently, there is a lack of trust and no agreement on scope, process or evidentiary standards between national payers or HTA bodies. Efforts should be made to harmonize standards and to build confidence before EFPIA can endorse assessments of relative efficacy (at launch) and relative effectiveness (over time) at the EU level.
6. Any assessment mechanism could facilitate a structured dialogue between HTA agencies and regulators during medicine development in order to coordinate, simplify and align their

requirements with regard to the clinical data and comparative assessments submitted by pharmaceutical companies.

7. The EFPIA HTA Task Force is empowered to, inter alia:
 - a. Explore different models for EU assessment of relative efficacy and effectiveness – how could this be done, where is it best done, over what timescale, etc.
 - b. Explore models to foster early dialogue during drug development between regulators, payers and other policy makers with patients, physicians and industry to align expectations and requirements.
 - c. Assimilate views of key stakeholders, including Member States.
 - d. Promote good practices in HTA (EFPIA Principles and implementation of High level Pharmaceutical Forum principles) and seek greater harmonization and trust between regulators, national payers and HTA bodies. This includes full support for EUnetHTA and the Joint Action on HTA. The dialogue should lead to a more consistent approach in reviewing the evidence base and should address issues related to clinical trial design and the generation of data post-approval.
 - e. Work, as appropriate, with EMA to shape the role of the Agency under the current mandate.

1.1 Definitions

Box 1 presents, as means of context, definitions for the key concepts referred to throughout the report.

Box 1. Definitions

Efficacy: The extent to which an intervention does more good than harm under ideal circumstances. (This comparison is usually based on data from clinical trials, and can be made at launch).

Relative efficacy: The extent to which an intervention does more good than harm compared to one or more alternative interventions under ideal circumstances. (This comparison is usually based on data from clinical trials, and can be made at launch).

Effectiveness: The extent to which an intervention does more good than harm under the usual circumstances of health care practice. (This comparison needs to be based on data collected in clinical practice over time).

Relative Effectiveness: 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. (This comparison needs to be based on data collected in clinical practice over time).

Cost effectiveness: Cost-effectiveness analysis (CEA) is a form of economic analysis that compares the relative costs and outcomes (effects) of two or more courses of action for a defined patient group and indication.

The EUnetHTA draws a distinction between generalisability and transferability, terms not clearly differentiated in the literature:

Generalisability refers to the extent as to whether the results of an HTA report can be readily adopted by other settings.

Transferability refers to the ability to apply information and/or data from one report for the user's target setting. The more context specific the data are, the less transferable they are.

1.2 Objectives and deliverables of the OHE Consulting Study

As part of this process, ten member companies of the EFPIA HTA Task Force commissioned OHE Consulting to carry out this work, with the ultimate objective of providing research data and analysis to enable EFPIA to support its policy position while contributing positively to the emerging dialogue around the development of HTA within Europe. This will also enable the pharmaceutical industry to work through how some form of a pan-European approach to the assessment of relative efficacy / relative effectiveness might work in practice in a way that could be acceptable both to Member States and to the pharmaceutical industry.

For this purpose, we have carried out the following steps:

1. To describe and critically review the current approach to the assessment of relative efficacy/relative effectiveness in Member States.
2. To identify and measure the underlying sources of variation of relative efficacy / effectiveness across Europe by reviewing the literature currently available on this question.
3. To look at the feasibility and desirability of pan-EU relative efficacy / effectiveness assessment as well as to give an early indication of its generalisability.
4. To consider how EFPIA could develop arguments in response to these findings that would advance its position on pan-European HTA given emerging pan-European HTA developments and engage constructively in dialogue around the scientific merit and public policy desirability of pan-European approaches.
5. To identify additional research that would enable the industry to develop a stronger evidence base to support its position on the practical issues involved in and the relative desirability of the options for harmonisation.

In assessing the current state of relative efficacy / effectiveness assessment in Member States we undertook to:

- Set out the current state of relative efficacy / effectiveness assessment in Member States in respect of methods and decision criteria to support pricing and reimbursement decisions for innovative medicines.
- Break down these approaches to identify the steps taken if any: in an ex-ante assessment of relative efficacy; a process of 'translation' of relative efficacy to determine relative effectiveness; some form of further 'translation' to a valuation of any added therapeutic benefit identified in the assessment of relative effectiveness.
- Seek to understand how these are conducted and applied to help clarify the journey that each country would need to take to adopt a pan-European assessment of relative efficacy / effectiveness and the scope and opportunity for harmonisation of elements of this technical approach.

The deliverables as set out in this report and the presentation we made to the companies are:

- Current use by the key Member States of relative efficacy and relative effectiveness data in their HTA and Pricing and Reimbursement processes.
- Current state of knowledge about the nature and causes of variations in treatment patterns, costs, prices and patient benefit across Europe'.
- The extent to which relative efficacy and relative effectiveness for an innovative medicine are likely to be similar between Member States and the extent to which adjustments may have to be made.
- Which countries were likely to find it straightforward to incorporate a pan-European approach to the assessment of either or both of relative efficacy and relative effectiveness and which were not.
- Implications of our findings for EFPIA's policy position and to underpin a dialogue with HTA stakeholders about the feasibility of a pan-European assessment of either or both of relative efficacy and relative effectiveness.
- The potential for conducting further research should the ten companies wish to commission this work. This would address, for example, gaps in understanding of the differences in the methods used by Member States to assess relative efficacy and of the extent of and causes of variations in relative effectiveness across Europe, and how the results of any further research can be used by EFPIA.

The research will help industry assess the implications and risks of moves toward pan-European HTA decision making, as well as informing debate around what is meant by 'contextualisation' and 'local differences' that need to be taken into account in HTA decision making.

1.3 Structure of the report

The structure of this report is as follows.

- Section 2 describes how eight EU countries use efficacy/effectiveness data within their health technology assessment (HTA) and/or pricing and reimbursement (P&R) mechanisms. The policy related question underpinning this strand of work is to understand how (some sort of) a pan-European assessment of relative efficacy/effectiveness might change the way these two variables are currently used by EU Member States.
- Section 3 reports on the current pan-European initiatives related to HTA. This includes EUnetHTA Joint Action, both in terms of the 2006-2008 Project and on the 2010-2012 Joint Action, which is still under discussion. Overall, Sections 2 and 3 provide an overview of the current environment on two fronts. First, on how Member States are using relative efficacy and relative effectiveness, and in particular, the differences and similarities across them. Second, on the policy environment at European level, driven by the joint working of the different HTA/P&R agencies themselves (and not only European). Understanding the current environment is essential to ascertain how the policy environment might evolve over the near future, and in particular, to understand which areas are more 'ready for harmonisation' at European level.
- Section 4 reports on the (systematic) literature review to establish whether there exist variations in relative effectiveness of drugs in any therapeutic area in two or more European countries, and in the case where heterogeneity was detected, to identify its underlying sources. The relevance of this section is twofold: (1) Differences in relative efficacy are assumed away – is this a problem? (2) Are differences in relative effectiveness a cause of concern?
- Section 5 pulls together the previous sections and puts the different pieces of the jigsaw together with an industry vision of the future.

2 Current Use of Effectiveness Data in Selected EU Countries

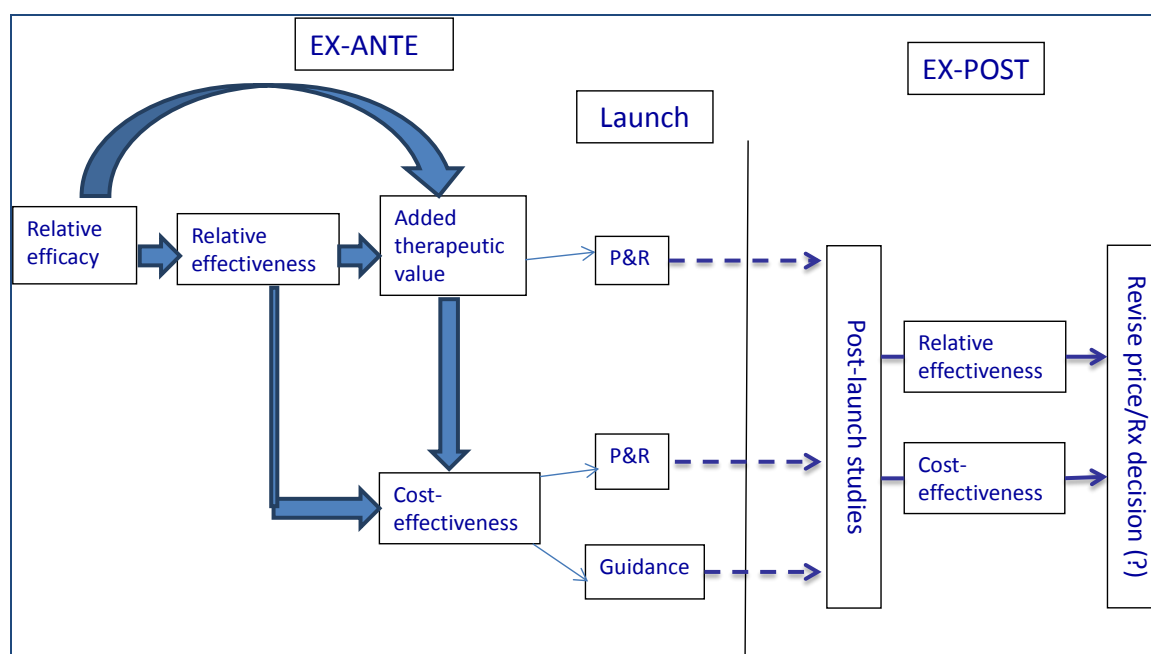
2.1 Linkages between key variables: framework

In this section we synthesise how eight EU countries use efficacy/effectiveness data within their health technology assessment (HTA) and/or pricing and reimbursement (P&R) mechanisms. Detailed accounts (where available) of how England and Wales, France, Germany, Italy the Netherlands, Scotland, Spain and Sweden use relative efficacy/effectiveness can be found in Appendix 1. These countries were chosen because of their size and/or because of the importance HTA has within their P&R mechanisms¹.

The information contained here is later used to help clarify the journey that Member States would need to take to adopt a pan-European assessment of either or both of relative efficacy and relative effectiveness and the scope and opportunity that exists for harmonisation of elements of any such assessments. The last sub-section in this chapter and Section 5 builds on this information.

As argued above, a key consideration is to understand the relationship between (relative) efficacy, relative effectiveness, measures of added therapeutic value/degree of innovativeness and cost effectiveness within HTA and P&R mechanisms, both at launch (i.e. ex-ante) and after launch (ex-post). Figure 2.1 outlines the possible inter-relations between the different variables, and shows all the possible transformations that can occur within any one country. It also shows the possibility of manufacturers needing to undertake post-launch studies.

Figure 2.1. Framework to explore linkages between key variables include post-launch studies



As shown by Figure 2.1, several transformations can occur prior to launch (under 'EX-ANTE'):

¹ When suggest that conducting similar analysis for other Member States is important in understanding the challenge of any pan-European approach.

- Efficacy to effectiveness
- Efficacy to added therapeutic value
- Effectiveness to added therapeutic value
- Effectiveness to cost effectiveness

Added therapeutic value, as argued below, has different meaning/dimensions in different countries; at this broad level, it refers to when countries use a formal way of classifying 'value' (as in the case of France and Italy). For the purpose of Figure 2.1, we have left this variable undefined, but we will comment on its different definitions below.

Based on these transformations (however done), the relevant body will then make a final decision regarding coverage (reimbursement), pricing and/or recommendations on use ('guidance') (under 'LAUNCH').

Figure 2.1 also represents a possible scenario after launch (under 'EX-POST'). Post-launch studies are becoming more important, as outlined below. Moreover, it is usually when relative effectiveness can become the key factor, especially if post-launch studies require companies to collect real-life data. There is also the possibility to undertake cost-effectiveness studies post-launch, using the relative effectiveness evidence collected. If such situation occurs, then the relevant authorities might revise the price and/or the reimbursement status of drugs, or alternatively, keep its price/reimbursement status as it is.

The purpose of this section is to identify the key issues underlying the different mechanisms/processes used by a selected number of Member States. For each country included here we adapt Figure 2.1 accordingly, and delete the arrows that are not relevant. This illustrates graphically the transformations that take place in that country and how these transformations are used. We now represent graphically each country in turn.

2.2 Linkages between key variables: country by country analysis

Figures 2.2 to 2.9 graphically represent how each country (in alphabetical order) transforms relative efficacy evidence into relative effectiveness, added therapeutic value and/or cost effectiveness (where applicable), and for what purpose. It also shows whether countries have an explicit mechanism for requesting and reviewing post-launch studies.

Figure 2.2. England and Wales

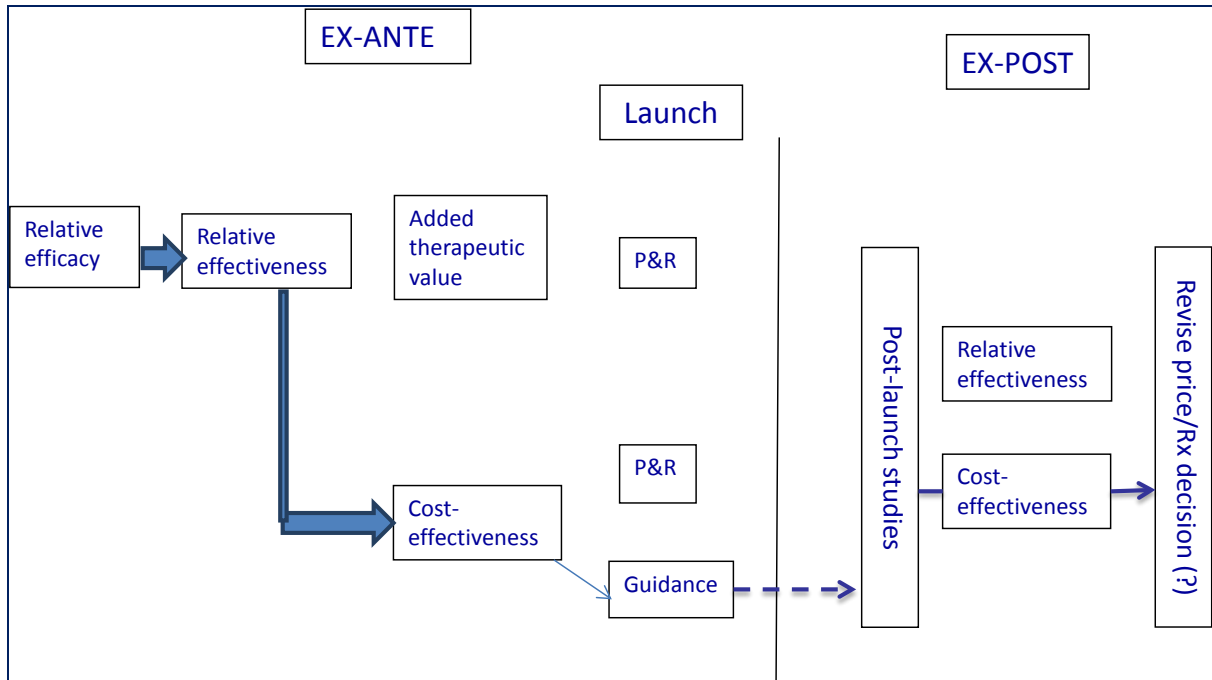


Figure 2.3. France

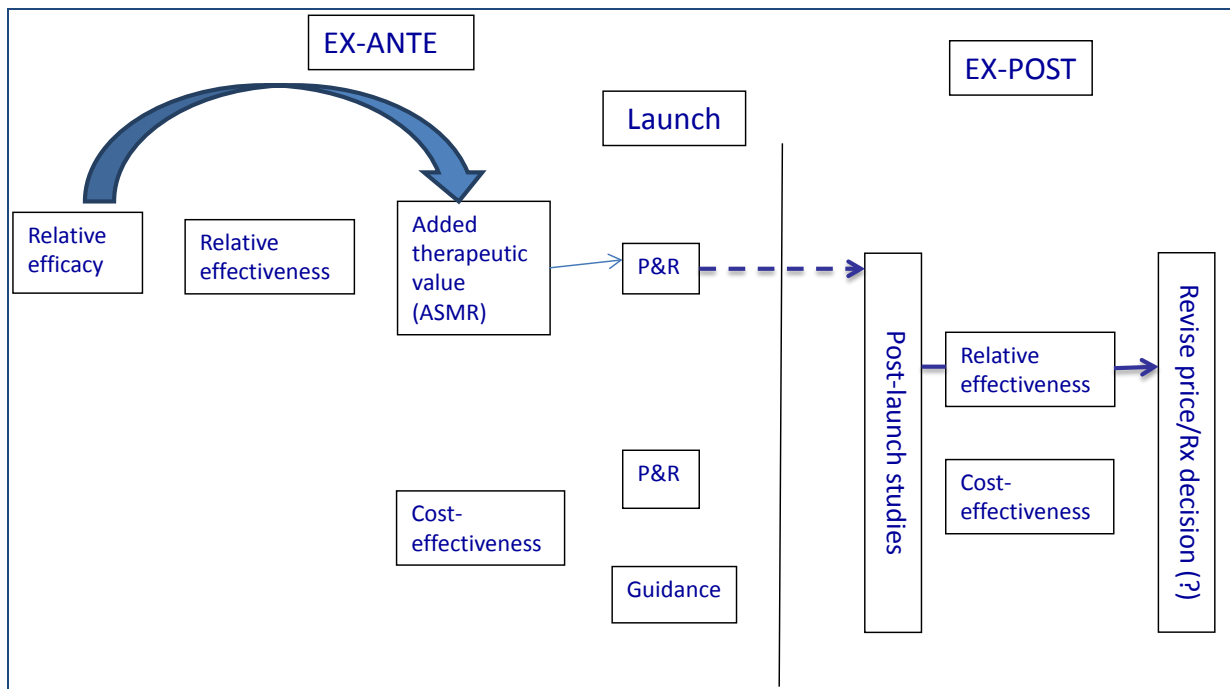


Figure 2.4. Germany

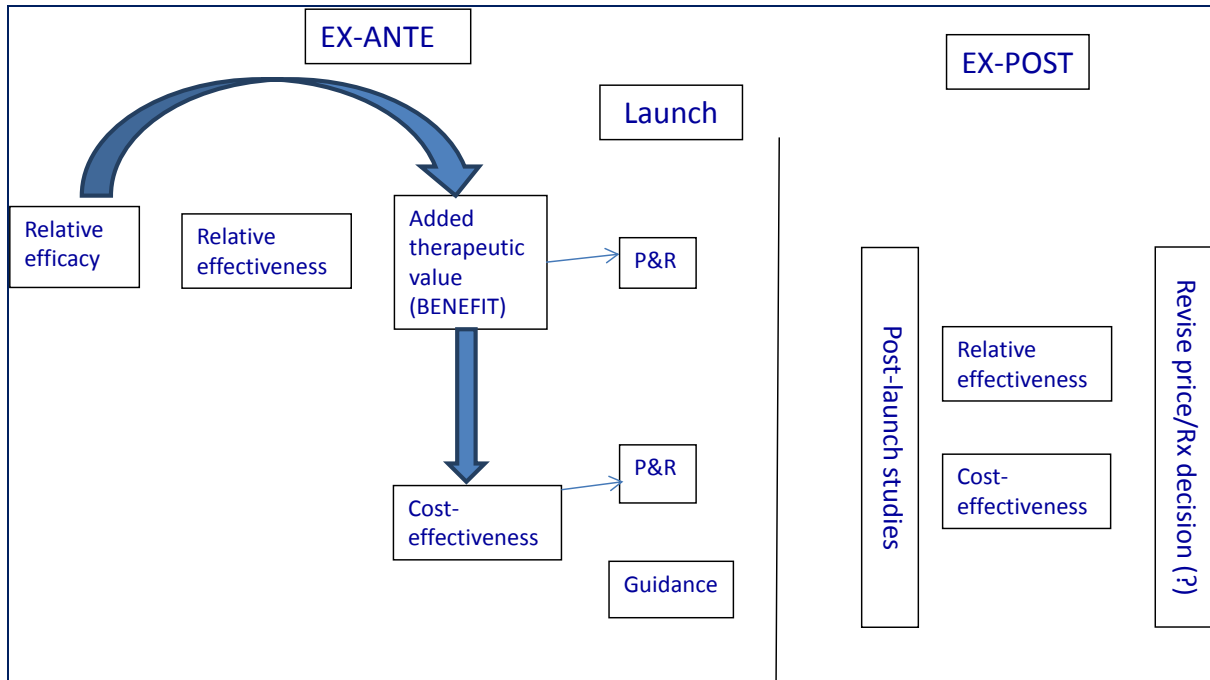


Figure 2.5. Italy

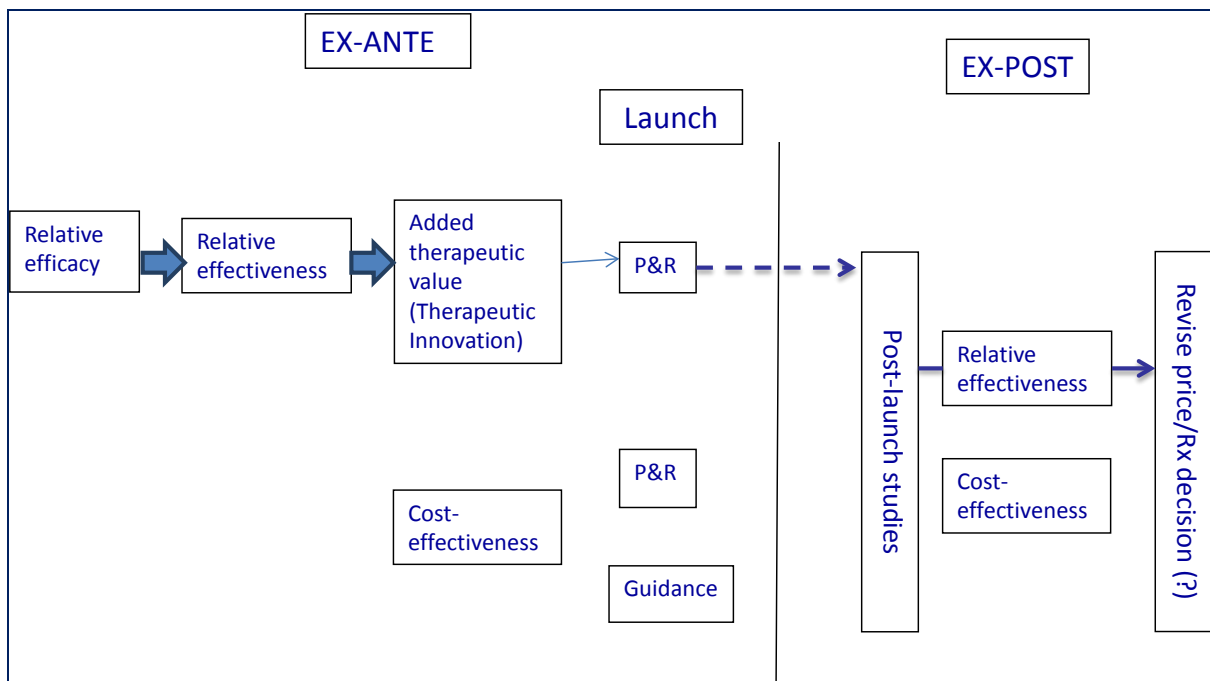


Figure 2.6. The Netherlands

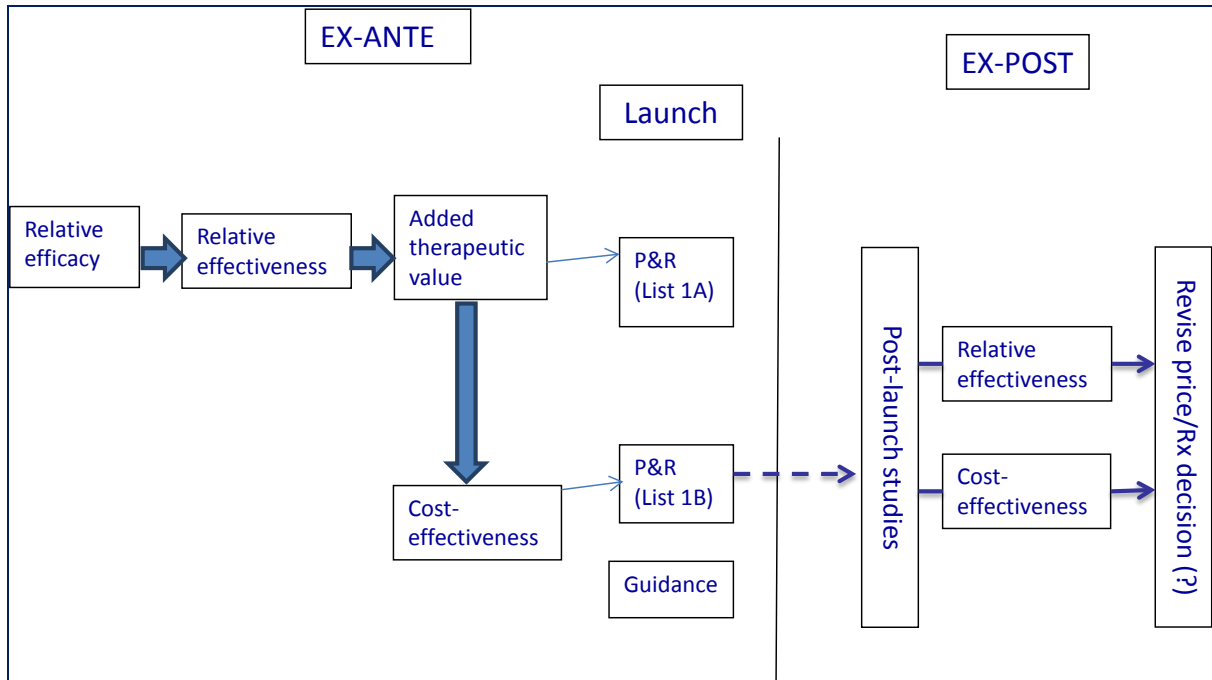


Figure 2.7. Scotland

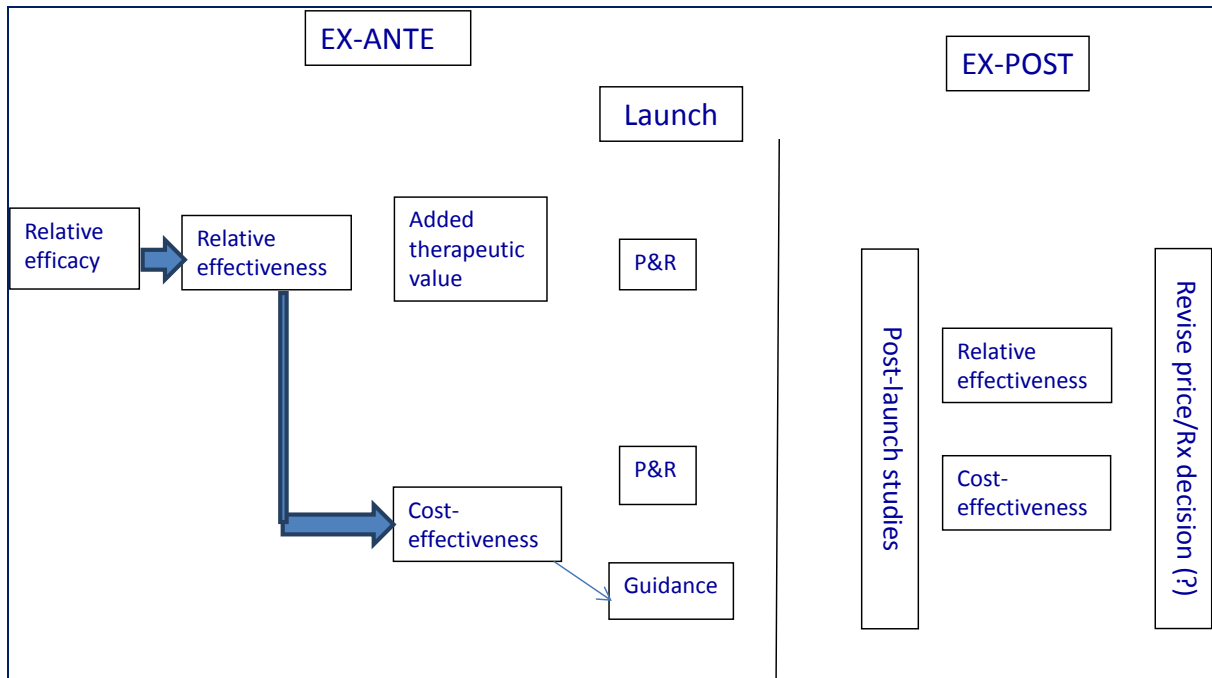


Figure 2.8. Spain

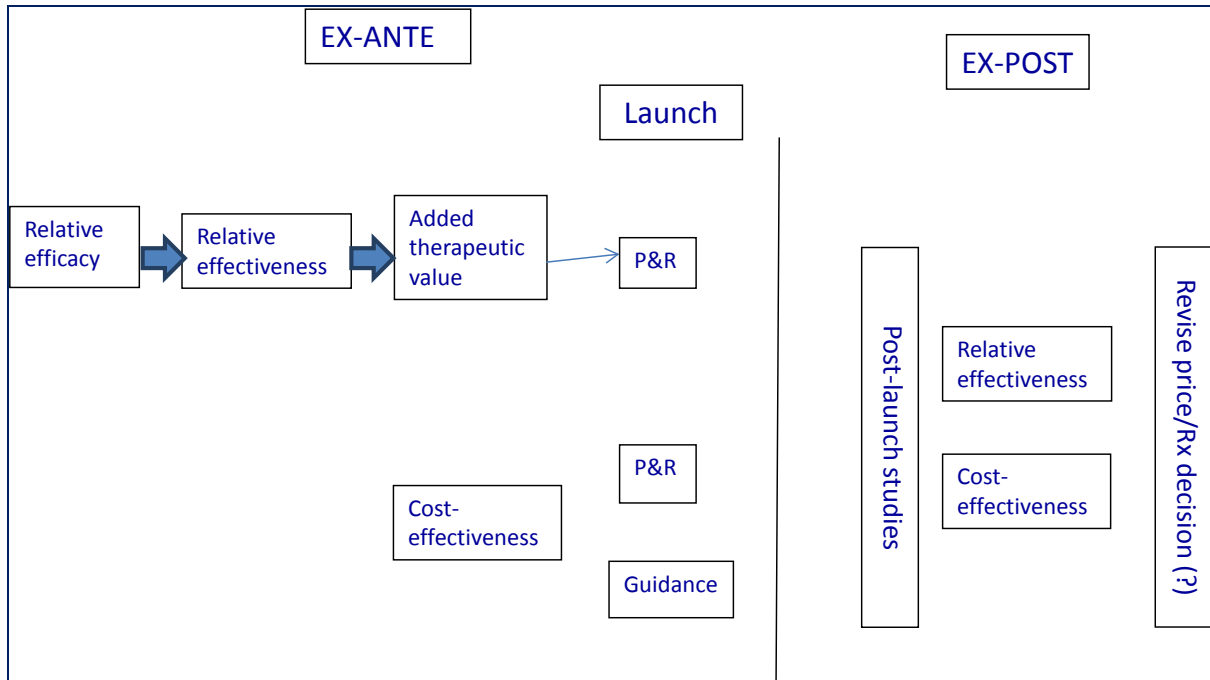
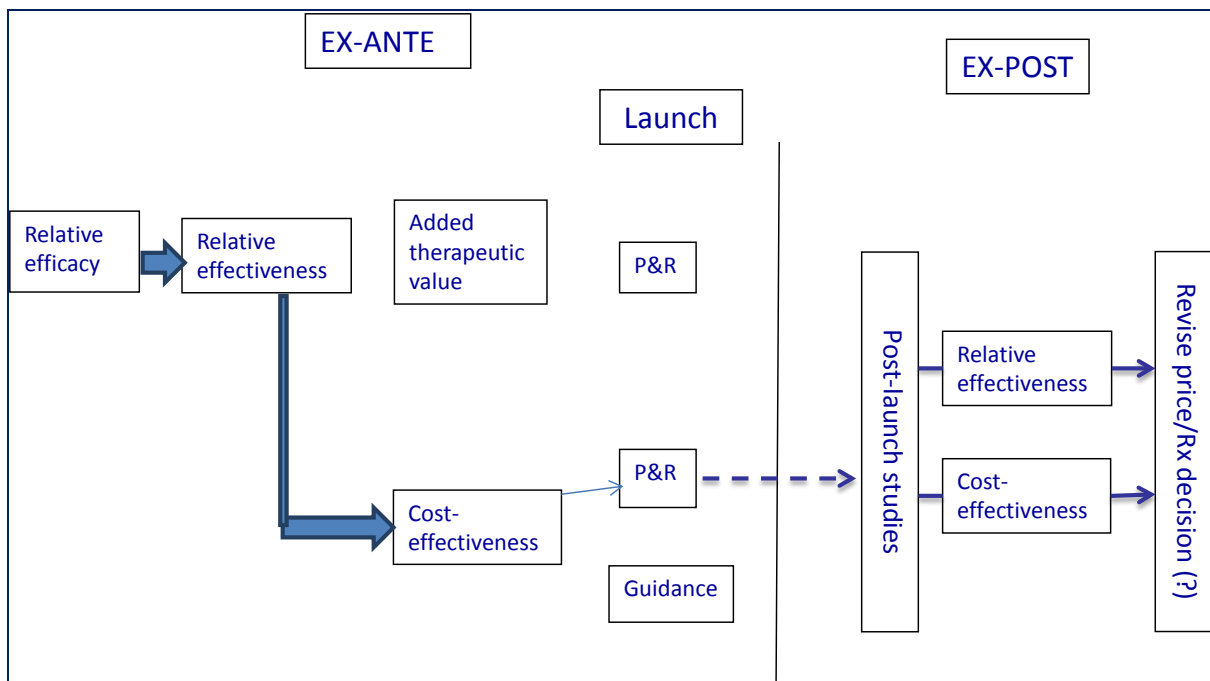


Figure 2.9. Sweden



P&R mechanisms may be classified several ways; many variables have to be taken into account when comparing the different systems. For the purpose of our work, we have decided to categorise pricing and reimbursement (P&R) systems into two main groups, based on whether or not they systematically use cost effectiveness/health economics:

1. Countries that use relative efficacy/effectiveness (based upon a systematic review/critical appraisal of the literature and/or clinical expert opinion) and budget impact.
2. Countries that use cost–effectiveness usually with one or both of budget impact and cost-effectiveness ‘thresholds’ – i.e. they are mainly ‘health economically driven’ systems.

From our selected list of eight countries, England and Wales, The Netherlands, Scotland and Sweden all have some relatively well established HTA mechanisms, in particular based on cost effectiveness. The remainder, France, Germany, Italy, and Spain do not. It is true, however, that France has a relatively well established system relying on classifying absolute and relative medical value, used to determine the reimbursement rate and prices of new medicines respectively. More recently, Italy has introduced a new innovation algorithm, and some regions in Spain use an algorithm to define the therapeutic value of new medicines. IQWIG in Germany, on the other hand, is still working on how it will carry out cost-effectiveness analysis despite its introduction a few years ago.

As shown by Figures 2.2-2.9, relative efficacy is the first variable looked at in all countries – and thus becomes key, either on its own or on how it is ‘converted’ into effectiveness and/or cost-effectiveness. There are then significant differences on how the process evolves. First, some countries seem to ‘jump’ from relative efficacy to some notion of added therapeutic value (in France, the ASMR, and in Germany, to benefit/additional benefit) that determines reimbursement status and prices.

Second, some countries try to translate efficacy data into effectiveness data to obtain some measure of added therapeutic value. This includes Italy’s recent innovation algorithm and Spain. These four countries, however, do not use cost-effectiveness currently. (Germany is starting to use cost-effectiveness, however, and we mention this below).

Third, some countries go one step further and use relative efficacy/relative effectiveness data to generate cost effectiveness evidence, and use the latter either for P&R or for guidance. Within these set of countries (England and Wales, The Netherlands, Scotland and Sweden), important differences are visible. Cost effectiveness in England and Wales and Scotland is used to offer guidance as prices are not negotiated. In Sweden cost effectiveness is used to determine reimbursement status, while in The Netherlands cost effectiveness becomes the key variable to determine P&R only for those medicines that are deemed to offer some additional therapeutic value. Note that while still in its early days, the new system being developed in Germany is similar in spirit to that of the Dutch, as only those drugs offering additional benefit will be subject to the ‘efficiency frontier’ approach.

Appendix 1 describes on a country-by-country basis (in alphabetical order) how relative effectiveness is used within HTA and P&R mechanisms. We have used a common template for each country, divided into two main parts. The first part discusses at a general level the P&R systems. This

presents a snapshot of the current environment – and some countries are currently engaged in further developments within their HTA system. For example, the UK recently announced a move towards a ‘value-based pricing’ system by 2014 (and they are in the process of implementing a Cancer Drugs Fund by 2011) that could have important implications for NICE and the SMC; IQWiG merits regular monitoring given its recent creation. We also understand that the situation will dramatically change from January 2011 with a mandatory relative effectiveness assessment within three months by IQWiG for each new drug, resulting in volume and pricing decisions without a hearing². The second part discusses at some detail some specific issues relating to how relative effectiveness is integrated into the different HTA/P&R system. As mentioned below, this is only possible for a sample of countries (i.e. those with guidelines). The common structure is as follows:

- Pricing and Reimbursement System: General Overview
- Guidelines
 - Analytical method - economic model
 - Health outcomes/QoL measure
 - Perspective
 - Evidence considered
 - Comparator
 - Choice of patient group
 - Analysis of uncertainty/sensitivity analysis
 - Costs
 - Time horizon
 - Discounting
 - Budget impact
 - Company/HTA organisation liaison
 - Presentation of methods/results

The main references used have been the (latest) guidelines/methodological reports published by the relevant bodies/organisations, where available. These guidelines mainly refer to those documents that outline, for instance, the process, methods and evidence used by the different bodies in their assessments. All references to guidelines are found in Appendix 1. However, not all countries have guidelines (at least in English). When this is the case, we have resorted to additional sources of information³, including a Pfizer internal survey of its European country affiliates.

² We thank Dr Frick for pointing this out.

³ Sorenson, C. (2009) The role of HTA in coverage and pricing decisions: A cross country comparison, *Euro Observer*, 11(1), 2-4; Sorenson, C., Drummond, M. and Kanavos, P. (2008) Ensuring value for money in health care: The role of HTA in the Europe Union., *Observatory Studies*, No. 11. Copenhagen: European Observatory on Health Systems and Policies.

Most health economics-driven countries have guidelines for economic evaluations. This is true for England and Wales, Scotland and Sweden. For The Netherlands, we are aware that guidelines exist, but we could not find them in English.

Germany's IQWiG has published a number of documents since its inception to describe its proposed methodology. While they are not guidelines per se (as for the HTA-driven countries), they contain relevant information for our purpose, as they describe the methodology used by IQWiG and states, for instance, IQWiG's preferences about the evidence they require.

We could not identify (official) guidelines for France, Italy or Spain. For France, however, the P&R process is relatively well understood. This is because the main components ('SMR' and 'ASMR' ratings) have been in place for many years. At the general level, the system is relatively clear – however, for some of the specific issues around relative effectiveness we did not find any official document that describes how these are incorporated in practice so we resorted to alternative sources.

Issues similar to those in the French system apply to Italy and Spain. The new system in Italy has been well described at a general level, but the details of how the new system is applied in practice are not known – at least we could not find any evidence of this in AIFA's webpage or other alternative sources.

In Spain, the general principles used to determine the price and reimbursement status of new medicines are known. But, again, the details are less certain, as the Spanish authorities have not published official guidelines. It is important to mention, however, that some Spanish regions are more active in the area of relative effectiveness, such as Catalonia, Andalusia and the Basque Country. These three regions have their own regional health technology evaluation committees. Given that the Department of Health (at national level) is still responsible for setting medicines' prices, these regional bodies aim to influence doctors' prescribing rather than negotiating medicines' prices directly.

We have synthesised the information contained in Appendix 1 in three tables. Table 2.1 gives a general overview of the relevant decision making body, the evidence required and the criteria for P&R decisions. It also includes a description on how separate (in time and in process) the relative effectiveness assessment and the funding/reimbursement decision are (under the heading 'Timing') and whether follow-up studies post-launch are mandated/recommended.

Tables 2.2 and 2.3 discuss more detailed dimensions. Table 2.2 compares some specific elements when relative efficacy / effectiveness is assessed, ranging from the analytical method to the criteria used to define the relevant comparator. Table 2.3 summarises the process used in these countries, including which stakeholders provide evidence and who does the assessment.

Cells with 'N/A' in all three tables signify that information was not available.

Before each table we provide a summary of the key issues that emerge from the information contained therein.

2.3 Pricing and reimbursement / market access: general overview

The column in Table 2.1 headed 'Role' describes the overarching objective of the relevant organisation. However, the decision making bodies have different specific roles as to three main variables of interest: coverage (reimbursement), pricing, and recommendations on use ('guidance').

The relevant body in most countries included in Table 2.1 (Germany, Italy, The Netherlands, Spain and Sweden) is in charge of both coverage and pricing. The exceptions include France, where coverage and pricing are separated as two different bodies are in charge of each function (under the auspices of HAS, however), and where the latter follows the former. In addition, NICE and the SMC are responsible for providing guidance/coverage – but do not negotiate prices.

From Table 2.1, it is clear that not all countries included use cost effectiveness analyses systematically to determine prices and/or reimbursement.

Three countries (France, Italy and Spain) use explicit definitions and classifications of therapeutic value, one based on efficacy (France) and two based on effectiveness (Italy and Spain – at least as far as we can tell based from our research). France has a long-standing tradition of classifying absolute (SMR) and relative medical value (ASMR) of medicines. The former is used for determining the reimbursement status while the latter is used to determine prices, together with other variables.

Italy has recently introduced a new 'algorithm' to classify the degree of therapeutic innovation of new products, based on three dimensions. Unfortunately, there is little publicly available information on how this algorithm is used in practice, and how it is used to set prices.

Some regions in Spain use decision trees to classify therapeutic improvements, although this is used to offer prescribing recommendations rather than setting prices (as this is done at national level). Moreover, not all regions use this decision tree.

England and Wales, The Netherlands, Sweden and Scotland rely on both clinical- and cost-effectiveness to determine pricing and reimbursement and/or to offer recommendations. NICE, the SMC and the TLV look at both clinical- and cost-effectiveness in parallel. The Netherlands, however, does this in two distinct phases. Therapeutic value is first used to determine a medicine's classification (either as List 1A or List 1B). For those medicines deemed to offer additional therapeutic value (List 1B), cost effectiveness is then used for price setting. On the other hand, medicines included in List 1B are clustered together with other therapeutic comparators, irrespectively of patent status.

IQWiG in Germany, on the other hand, is still working on how it will carry out cost-effectiveness analysis – but it follows an approach similar to The Netherlands, in that a health economic evaluation is carried out only for those technologies with additional benefit. However, their methodology is unique in that they use efficiency frontiers to set maximum reimbursement prices.

Only two countries from our list (England and Wales, and Scotland) explicitly have a threshold range as a criterion for their decisions. However, exceptions are allowed to deviate from this threshold.

In most countries the relative effectiveness assessment feeds into the reimbursement/funding and pricing decision. This is because in these countries, with the exception of Germany and Sweden, a (confidential) negotiation stage occurs that involves third party payer and the manufacturer. The TLV, on the other hand, has stated publicly that they do not negotiate prices. England and Wales and Scotland do not have what could be considered a normal 'pricing and reimbursement' system. Manufacturers enjoy 'free pricing' (with some constraints) and can launch their products after obtaining marketing authorisation. NICE and SMC offer recommendations as to whether their respective NHS should use the medicines – but they do not negotiate prices. That is why under the last column we have 'Same time' for these two jurisdictions.

An important dimension to consider is whether the different HTA / P&R bodies can recommend and/or impose follow-up studies as part of their decision making process, in a similar spirit to the concept of 'coverage with evidence development'. Unfortunately, few guidelines discuss this. Still, some countries have used some form of conditional reimbursement (Italy, The Netherlands and Sweden) with the condition that companies collect real life data to demonstrate medicines' clinical- and cost-effectiveness. This is also the case in France. Unfortunately, limited evidence as to how this is done in practice is available for many of these countries. An important policy initiative was recently introduced in England and Wales, under the 2009 Pharmaceutical Price Regulation Scheme (PPRS): the possibility for flexible pricing and patient access schemes (PAS). The objectives of both schemes are to increase understanding of the underlying cost-effectiveness of the technology itself by collecting evidence alongside use. The PASs to date have increased access to medicines, but no schemes proposed to date have helped to tackle outcomes uncertainty.

Overall, HTA driven countries using clinical- and cost-effectiveness data are particularly interested in relative effectiveness. For non-HTA driven countries, and in particular Germany, relative efficacy demonstrated by RCTs seems to be one of the key variables and, as such, are less interested in how relative effectiveness is modelled based on efficacy.

Table 2.1 Pricing and Reimbursement / Market Access – General Overview

	Decision making body	Role	Evidence required	Criteria for P&R decisions	Timing between relative efficacy or relative effectiveness assessment and Rx/funding decision	Follow-up studies
England and Wales	NICE	NICE provides guidance, sets quality standards and manages a national database to improve people's health and prevent and treat ill health	Clinical and cost effectiveness	<ul style="list-style-type: none"> • Threshold of £20-30,000/ QALY • Special consideration for end of life treatments targeting small populations • Other factors on a case by case basis 	Same time	<ul style="list-style-type: none"> • NICE offers recommendations for future research topics, based on evidence gaps (but no consideration for funding source) • Flexible Pricing and Patient Access Schemes
France	HAS	To improve quality of care	Clinical therapeutic / medical value	<ul style="list-style-type: none"> • Absolute medical value (SMR): Major – Insufficient (3 levels) • Incremental medical value (ASMR) based on clinical efficacy: Major – No improvement (5 levels) 	Assessment (SMR/ASMR) feeds into Rx/funding decision	<ul style="list-style-type: none"> • Price-volume agreements • Post-marketing studies (observational studies and risk-sharing)
Germany	IQWiG	To evaluate the benefits, harms, and economic implications of interventions to contribute to the continuous improvement in the quality and efficiency of health care in Germany	<ol style="list-style-type: none"> 1. Efficacy data 2. Health economic evaluation for technologies with additional benefit versus already available technologies 	IQWiG will use efficiency frontiers to set maximum reimbursement price (only if there are alternative treatment(s))	Varies with process	N/A

	Decision making body	Role	Evidence required	Criteria for P&R decisions	Timing between Rel efficacy/effectiveness assessment and Rx/funding decision	Follow-up studies
Italy	AIFA	Set fair pharmaceutical policies and assure their consistent application nationwide, manage value and cost of medicines and promote pharmaceutical R&D	Clinical therapeutic value and innovation	<ul style="list-style-type: none"> • Disease seriousness • Availability of treatments • Therapeutic effect 	Assessment (innovation algorithm) feeds into Rx/funding decision	Use of conditional P&R (risk-sharing and payment for performance)
The Netherlands	CVZ	The CVZ has an important part in maintaining the quality, accessibility and affordability of the Dutch health system	Clinical and cost effectiveness (latter for selected products as a second step)	<ol style="list-style-type: none"> 1. Therapeutic value (List 1A/ List 1B) 2. Health economic evaluation for 'innovative' technologies (List 1B only) 	Assessment feeds into Rx/funding decision (List 1A/List 1B)	Conditional reimbursement for selected drugs (three years)
Scotland	SMC	To provide advice across Scotland about the status of all new medicines, including new formulations and indications (licensed from January 2002)	Clinical and cost effectiveness	<ul style="list-style-type: none"> • Threshold of £20-30,000/ QALY • Other factors ('modifiers') on a case by case basis 	Same time	N/A

	Decision making body	Role	Evidence required	Criteria for P&R decisions	Timing between relative efficacy or relative effectiveness assessment and Rx/funding decision	Follow-up studies
Spain	Ministry of Health (General Directorate for Pharmacy and Medical Products)	Responsible for the coordination of pharmaceutical policy and pharmaceuticals financing	Clinical therapeutic value	<ul style="list-style-type: none"> • Seriousness • Needs of certain groups • Therapeutic and social utility • Limits of drugs expenditure • Alternative treatments • Degree of innovation 	Assessment feeds into Rx/funding decision	N/A
Sweden	TLV	To determine whether a pharmaceutical product or dental care procedure should be subsidised by the state	Clinical and cost effectiveness	Three principles: <ol style="list-style-type: none"> 1. Human value 2. Solidarity 3. Cost-effectiveness 	Assessment feeds into Rx/funding decision	Reimburse with conditions (including company supplying observational study with supplementary information on the clinical effect and cost-effectiveness)

2.4 Methods of relative effective / effectiveness assessment

Decision making bodies have not published explicit definitions of relative efficacy and/or relative effectiveness in their guidelines. Some of them, however, list the different criteria used to classify the key 'metrics', as shown in Table 2.2. For instance, France lists the factors considered when appraising the SMR and ASMR rating. For the former, efficacy and severity of the disease seem to be the key two components. The SMR rating defines the absolute medical value, irrespectively of the competitors that may already be available in France. The ASMR component, however, is related directly to relative effectiveness as looks at the relative medical value compared to existing treatments. For instance, a generic drug might be deemed to have the highest possible SMR rating, but will be classified as having no additional medical value. The evidence published by HAS summarising their SMR/ASMR decisions clearly shows that only a very limited number of new drugs have been deemed to offer significant relative value, although most drugs have a major absolute value. To our knowledge, however, there is no explicit quantitative definition of 'significant'.

For France and Italy, it seems that similar dimensions are used in their definition of 'therapeutic innovation' (which could be considered as an analogous term for relative effectiveness). The availability of existing alternatives and what the new drug offers in terms of therapeutic effect are important dimensions that affect prices in these two countries. These two countries seem to use a similar approach irrespectively of the therapeutic area. This also can be said of those countries that are HTA-driven. Germany, the exception to this, explicitly argues that each efficiency frontier is therapy-driven.

There is a general consensus within most HTA-driven countries about the analytical method (cost utility analysis) and the health outcomes (QALYs) preferred. While most of these HTA bodies comment in their guidelines on the weaknesses of QALYs, they still strongly recommend that companies use them in modelling – and provide good evidence for not using them.

These HTA driven countries also request companies to be very explicit about their modelling method, in particular about assumptions and data inputs. These countries acknowledge the need to model (clinical- and cost-) effectiveness based on clinical evidence – and for that purpose all available evidence needs to be collected and synthesised in a robust and transparent manner.

For this purpose, these countries also tend to be more flexible about the evidence considered. While all HTA bodies express a strong preference for head-to-head RCTs, they also acknowledge that other sources of evidence need to be used to supplement RCT data. Indeed, NICE argues that in considering cost effectiveness, evidence from other study designs, in addition to that from RCTs, will be necessary. Moreover, while evidence on cost effectiveness may be obtained from new analyses, a systematic review of published, relevant evidence on the cost effectiveness of the technology also should be conducted. IQWiG, on the other hand, is more critical about the use of non-RCT data. It argues that RCTs are usually possible and practically feasible in the assessment of drugs and that the use of non-randomised intervention studies or observational studies is justified only in exceptional cases.

The criteria used to select the comparator tend to be similar in spirit in all countries that make an explicit comment about the choice of comparator for the modelling: therapies routinely used, usual

treatment, therapy most likely to be replaced, etc. However, the important element underpinning the choice of the comparator is that each country always refers to its own local setting i.e. it is the 'appropriate' local comparator, including those that have already been previously recommended/accepted (e.g. NICE). This is consistent with the request of HTA bodies to ask for evidence that reflects local conditions; if the efficacy data needs to be modelled because of lack of 'local' evidence, companies need to show clearly what assumptions have been used for this purpose. The SMC, for instance, is explicit about this, and states that the appropriate comparator for the cost effectiveness modelling may be different to the comparator in the clinical studies programme for the medicine, but, if so, the manufacturer must then carry out an indirect comparison.

The use of surrogate markers also is discussed by some of the HTA bodies. Again, there is some consistency across them in that manufacturers are asked to provide good and robust evidence showing the relationship between surrogate points and relevant end points.

In terms of the criteria used to decide on whether a new medicine is clinically superior to its comparator, the relative size of the clinical effect seems to be the dominant one, followed by the absolute size of the clinical effect.

A very important issue here is what the relevant guidelines state about sub-group analysis. This is increasingly important, especially as many bodies, such as NICE and the TLV, offer restricted recommendations (i.e. reimbursement is limited to a certain sub-group of patients because cost effectiveness differs between different population sub-groups). TLV's guidelines, for example, explicitly mention this. They state that separate calculations must be made for different patient groups where the treatment is expected to have different cost-effectiveness (e.g. separately for men and women in different ages and with differing degrees of severity for the illness/symptom or with different risk levels). The guidelines also state that if the results from randomised clinical studies contain only a portion of the patient population to which the application refers, modelling should be undertaken to illustrate cost-effectiveness in the remaining patient groups. However most guidelines do not address sub-group analysis and we have not included this issue in Table 2.2.

Table 2.2. Methods of Relative Efficacy / Effectiveness Assessment

	Definition of relative effectiveness	Analytical method	Health outcomes	Evidence considered	Perspective (Outcomes)	Comparator	Uncertainty	Surrogate markers	Criteria to define 'clinical superiority' vs. comparator
England and Wales	Clinical effectiveness: The extent to which an intervention produces an overall health benefit, taking into account beneficial and adverse effects, in routine clinical practice	Cost effectiveness (CUA preferred)	QALYs	<ul style="list-style-type: none"> • Strong preference for 'head to head' RCTs • But others accepted (non-RCTs; indirect comparisons; mixed treatment comparison (MTC); meta-analysis) 	All health effects on individuals	Therapies routinely used (including current best practice)	Full account of assumptions and data inputs. Sensitivity analysis required.	N/A	<ul style="list-style-type: none"> • Relative size of clinical effect • QALYs
France	SMR: efficacy/ tolerance; severity of the disease; existence of therapeutic alternatives; place in the therapeutic strategy (first line, second line, etc.); and public health impact ASMR: Improvement (therapeutic efficacy; tolerance)	N/A	Final outcomes preferred: mortality, morbidity and quality of life	Clinical studies (head-to-head RCT preferred); systematic literature review and synthesis	N/A	Approved, listed pharmaceuticals of same therapeutic category, in terms of: <ul style="list-style-type: none"> • used most regularly (by treatment days) • with cheapest treatment costs • included in positive list most recently 	N/A	N/A	N/A

	Definition of relative effectiveness	Analytical method	Health outcomes	Evidence considered	Perspective (Outcomes)	Comparator	Uncertainty	Surrogate markers	Criteria to define 'clinical superiority' vs. comparator
Germany	Relative effectiveness is not mentioned in the IQWiG methods paper - reference made only to benefit and additional benefit	Efficiency frontier per therapeutic area	Primary clinical measures: mortality, morbidity, HRQoL and validated surrogates	RCTs and meta-analyses of individual RCT results are mandatory. Use of non-randomised interventions studies or observational studies justified only in exceptional cases.	SHI insurants	Most used drugs, but assessment looks at indications rather than just individual drugs	Impact of uncertainty on model results has to be investigated by means of sensitivity analyses	Need to prove the association between the change in the surrogate outcome and the change in the patient-relevant outcome	<ul style="list-style-type: none"> • Absolute size of clinical effect • Relative size of clinical effect
Italy	N/A	Innovation algorithm: Therapeutic innovation = disease seriousness + treatment availability + therapeutic effect	N/A	N/A	N/A	N/A	N/A	N/A	N/A

	Definition of relative effectiveness	Analytical method	Health outcomes	Evidence considered	Perspective (Outcomes)	Comparator	Uncertainty	Surrogate markers	Criteria to define 'clinical superiority' vs. comparator
The Netherlands	Therapeutic value: incremental efficacy, effectiveness, quality of life and safety (including fewer/lesser side effects and greater tolerability)	Cost utility and/or cost effectiveness analysis	QALYs	Modelling from efficacy to effectiveness allowed (with full account of assumptions. Meta-analyses are preferred but observational studies and MTC is accepted)	Societal	The standard or usual treatment (first choice for which effectiveness has been proven) and alternatively the most used technology in the particular indication	Sensitivity analysis required	N/A	<ul style="list-style-type: none"> •Absolute size of clinical effect •Relative size of clinical effect
Scotland	N/A	Cost effectiveness (CUA preferred)	QALYs	RCTs, but others accepted (non-RCTs; indirect comparisons; meta-analysis)	All health effects on individuals	Most likely to be replaced if the medicine under consideration is accepted	Sensitivity analysis required to check the robustness of the ICERs and under which circumstances the ICER exceeds £20,000 and £30,000	Companies need to provide details of any association between surrogate markers and health benefits or disadvantages to patients	<ul style="list-style-type: none"> •Relative size of clinical effect •QALYs
Spain	No explicit definition	Cost effectiveness analysis	N/A	Individual RCTs preferred; meta-analyses of RCTs, observational studies, modelling of effectiveness accepted	Payer	Most used and/or cheapest.	N/A	N/A	<ul style="list-style-type: none"> •Relative size of clinical effect •Expected disease management impact

	Definition of relative effectiveness	Analytical method	Health outcomes	Evidence considered	Perspective (Outcomes)	Comparator	Uncertainty	Surrogate markers	Criteria to define 'clinical superiority' vs. comparator
Sweden	As stated in The Law of Pharmaceutical Benefits (2002:160): there are no other available pharmaceuticals or treatments that according to such a balance between intended effect and side effect could be considered as significantly more appropriate	Cost effectiveness analysis	QALYs	RCTs; indirect comparisons. Modelling of effectiveness is allowed and meta-analyses of individual RCTs are mandatory	Societal	Most appropriate alternative treatment in Sweden (e.g. the most used)	Sensitivity analysis of central assumptions and parameters required	If surrogate end-points are used, the submission should also include modelling assumptions to illustrate the effects on mortality and morbidity, i.e. QALY's gained	<ul style="list-style-type: none"> •Relative size of clinical effect •Ethical considerations

2.5 Process of relative efficacy / effectiveness assessment

In most cases, the onus to provide the evidence and the cost effectiveness modelling lies on the manufacturer (NICE's STA, SMC). The role of the HTA body is then to evaluate that evidence and to provide a critique of the model. Only for NICE's MTA and IQWiG is additional modelling carried out and compared to the manufacturer's submission.

To our knowledge, external groups are involved in the assessments only in England and Wales (the so-called 'Assessment Group' for MTAs). IQWiG also states that external expertise is used where appropriate when producing Reports and Rapid Reports. The exact details of this involvement are not stated, however.

The countries that discuss stakeholder involvement in their guidelines provide, at least on paper, the opportunity for the public, including manufacturers, to provide input at different points in time during the process; from the initial stages (Scoping for NICE and Hearing Procedure [which is not a scoping process per se such as the one in NICE] in Germany) to when the preliminary recommendations are published. However, what is not clear is the weight given in making the final decision to the different inputs provided throughout the process.

NICE and IQWiG are relatively clear in outlining the different steps in the process. Both follow a similar process in that there is some preliminary discussion to define the scope of the appraisal, the assessment is then carried out, a preliminary report is published, and then the final decision is made. The preliminary and final decisions are made public in the process.

Table 2.3. Process of Relative Effectiveness Assessment

	Who provides the evidence	Who does the assessment	Stakeholder involvement	Steps of process: what gets documented and what is available
England and Wales - MTA	Assessment Group (independent); Manufacturers and sponsors; patient/carer groups; healthcare professionals; clinical experts	Assessment Group; Manufacturers and sponsors	<ul style="list-style-type: none"> Stakeholders (manufacturers, patient/carer groups, healthcare professionals and clinical specialists) take part in Scoping phase Consultation period between Appraisal Consultation Document (ACD) and Final Appraisal Determination (FAD) 	<ul style="list-style-type: none"> Scoping Assessment Appraisal
England and Wales - STA	Evidence Review Group (independent); Manufacturers and sponsors; patient/carer groups; healthcare professionals; clinical experts	Manufacturers and sponsors (mainly); ERG may undertake additional analysis	<ul style="list-style-type: none"> Stakeholders (manufacturers, patient/carer groups, healthcare professionals and clinical specialists) take part in Scoping phase Consultation period between Appraisal Consultation Document (ACD) and Final Appraisal Determination (FAD) 	<ul style="list-style-type: none"> Scoping Assessment Appraisal
France	Manufacturers	Transparency Commission (TC); CEPS	N/A	<ul style="list-style-type: none"> TC: determines SMR/ASMR status CEPS: price negotiations

	Who provides the evidence	Who does the assessment	Stakeholder involvement	Steps of process: what gets documented and what is available
Germany	Manufacturers; IQWiG	IQWiG	<p>IQWiG Reports:</p> <ul style="list-style-type: none"> •Public (including companies) invited to submit written comments (hearing procedure) on preliminary report plan (published) •Public can submit comments after Preliminary Report is published 	<p>Reports:</p> <ul style="list-style-type: none"> •Work is commissioned •Report Plan •Preliminary Report •Final Report <p>NB Process is faster, with no hearing procedure, for 'Rapid Reports'</p>
Italy	N/A	N/A	N/A	N/A
The Netherlands	Manufacturers	CVZ	N/A	N/A
Scotland	Manufacturers	Manufacturers	No face to face contact between manufacturers and reviewers. But possibility of interaction, for clarification, via the Secretariat	N/A
Spain	N/A	Manufacturers	N/A	N/A
Sweden	Manufacturers	Manufacturers	Regular contact between company and government office, including face to face meetings	N/A

2.6 Implications for a pan-European assessment (of relative efficacy)

As noted above, we have analysed how different Member States use relative effectiveness evidence for several reasons. The first is to provide context for exploring the impact that some form of a pan-European assessment might have on the current situation. And, second, to explore which areas might be more or less amenable for harmonisation.

Based on our desk research, it seems to us that:

- While there appears to be commonality around the starting point of relative efficacy, based on RCTs, it is unclear whether the methods used to identify, include, and analyse RCTs are similar. It is also unclear to what extent the comparators used in any assessment of relative efficacy might differ;
- The use made of a relative efficacy assessment seems to be quite different. Some countries are using this to make an innovation assessment. This is true, for example, in France and Italy. Others use it as a starting point for assessing relative effectiveness, but here the approaches are quite different. For example, Germany talks about benefit and additional benefit, rather than relative effectiveness. And HTA-driven countries use relative efficacy to model clinical- and cost-effectiveness. While there are some commonalities in the different approaches, how the relative effectiveness assessment is used can be different: either to determine prices and reimbursement, or to issue recommendations on which drugs to use, and how (for instance, whether restricted to a certain patient subpopulation). Moreover, other dimensions, other than clinical and cost-effectiveness, are important in the decision making process – for example, whether and how many treatments are available and disease seriousness.

3 An Overview of Pan-European Initiatives Related to HTA

3.1 The policy context in Europe

The G10 High Level Group report in 2002 established the EU agenda in this area. It focused, inter alia, on the need to 'encourage the development and quality of HTA'; and to 'share national experiences and data while recognising that relative evaluation should remain a responsibility of Member States'.

More recently the High Level Pharmaceutical Forum (HLPF), created in order to take forward some of the G10 recommendations, has reported. The HLPF had an initial mandate 'to discuss the competitiveness of the European pharmaceutical industry and related public health considerations, with a specific focus on information to patients on disease and treatment options, relative effectiveness assessments and pricing and reimbursement of medicinal products'. Working Groups on Pricing and Relative Effectiveness, established as part of that Forum, recommended information exchanges, enhanced HTA quality and timely access to innovative medicines.

The HLPF brought consensus to defining both relative efficacy and relative effectiveness terms. The former refers to the extent to which an intervention does more good than harm, under *ideal* circumstances, compared to one or more alternative interventions. The latter refers to the extent to which an intervention does more good than harm, compared to one or more alternative interventions for achieving the desired results when provided under the *usual* circumstances of health care practice. This leaves unanswered the extent to which relative effectiveness refers only to clinical effects, or might also include wider notions of benefit and value.

In their final conclusions on relative effectiveness, the HLPF acknowledged 'the distinction between the scientific assessment of the relative effectiveness of medicinal products and health-economic assessments of their costs and benefits'. The HLPF also endorsed 'the aim of relative effectiveness assessment to compare healthcare interventions in daily practice and classifying them according to their added therapeutic value'. Final recommendations on implementation of good practice principles and the exchange of information on the relative effectiveness were made to improve the data availability and transferability.

There are already a number of possibilities for a pan-European approach to assessing relative efficacy and relative effectiveness. The EMA has indicated an interest in pan-European relative efficacy. An existing HTA network, MEDEV, has indicated an interest in pan-European assessments of relative effectiveness. The European Commission is seeking implementation of the results of the EUnetHTA Project 2006-2008 and the Pharmaceutical Forum, through the Joint Action on HTA, is funding additional work by EUnetHTA to develop further pan-European HTA initiatives and emergent collaborations between HTA bodies. It is possible to envisage a 'mutual recognition' possibility arising in which the agencies comprising EUnetHTA experiment with different agencies taking the lead in preparing a pan-European relative assessment dossier (of efficacy in the first instance) for particular innovative medicines. It may well be that such an initiative would not need legislation at either European or national levels, at least initially.

The next sections report more detail on the work from EUnetHTA and from other multilateral collaborations to address the assessment of relative effectiveness in Europe, as the tasks undertaken by the Swedish EU presidency initiative.

3.2 The European Network for Health Technology Assessment (EUnetHTA)

The EUnetHTA Project (2006-2008) took account of the previous European collaborative work on HTA and established an organisation that sought inclusion of all Member States, with wide involvement of experts, involving 64 partners (50 from European Countries, five from Australia, Canada, Israel and USA and nine international organisations) in eight Work Packages (WP1-WP8).

The *strategic objectives* of the EUnetHTA Project were to:

1. Reduce overlap and duplication of effort and hence promote more effective use of resources in the HTA process;
2. Increase HTA input to decision-making in Member States and the EU and hence to increase the impact of HTA;
3. Strengthen the link between HTA and health care policy making in the EU and its Member States; and
4. Support countries with limited experience with HTA.

In order to reach these objectives, additional *specific objectives* were defined to facilitate rapid, productive collaboration that would lead to the development of a range of practical tools to deliver the strategic objectives. Eight Work Packages (WP) were aligned with specific objectives and each was expected to produce specific deliverables.

3.2.1 Results from the EUnetHTA Project 2006-2008

The work of the EUnetHTA Project during the three year period involved two clear strands: 1) delivering tools and information to support HTA in Europe, and 2) developing a well-functioning network of national HTA organisations that could share information and undertake joint work.

Practical tools and systems to support the development of HTA information were created and articulated through three different Work Packages: WP4 HTA Core Model, WP5 HTA Adaptation Toolkit, and WP7 Monitoring development of emerging and new technologies and prioritisation of HTA. The remaining WPs were orientated to develop the specific objectives in the second strand: WP1 Coordination, WP2 Communications, WP3 Internal evaluation, WP6 Transferability of HTA to health policy, and WP8 System to support HTA in Member States with limited institutionalisation of HTA.

The WP5 HTA Adaptation Toolkit (EUnetHTA 2007a) developed a series of checklists providing a systematic method to determine the policy relevance, and the reliability and transferability of data and information, in an existing report to a new context. It is designed to help the user to determine whether any existing HTA report (or part of it) addresses similar issues, is of sufficient quality and is applicable to the new context. The checklists cover five main domains: technology use, safety,

effectiveness, economic evaluation, and organisational aspects. An interactive version of the Toolkit was planned to be available by 2009.

The WP7 aimed to provide relevant and structured information on new and emerging technologies and tools to monitor their development. First, a newsletter series 'On the Horizon' were designed to provide this information and circulated to the HTA agencies and policy makers. Secondly, a web-based toolkit was created to be used by HTA organisations to enquire about on-going work or to share existing work on new health technologies. This is a database and structured questionnaires providing information about the level of diffusion of the technology in different healthcare systems, the status of any HTA, monitoring actions including protocols and results and use of new evidence for a final reimbursement/coverage decision. This project aimed to facilitate joint working to generate evidence across HTAs agencies.

Other areas of work developed by the EUnetHTA Project focused on developing a well-functioning network of national HTA organisations and taking on board all relevant stakeholders (in the Stakeholder Open Forum run by the WP6) to share information and gain feedback from them. The WP8 produced handbooks and reports to share the HTA knowledge across Europe and help countries with limited institutionalisation of HTA.

3.2.2 The HTA Core Model

The HTA Core Model probably is one of the most relevant contributions of the EUnetHTA Project despite the need for its future improvement and applicability. The work was led by the Finnish Office for Health Technology Assessment (FINOHTA). Two applications of the model for the most commonly assessed health technologies and one pilot for each of them were developed: 1) the medical/surgical interventions model (EUnetHTA 2007b) was tested with a Core HTA for drug eluting stents (DES) vs. bare metal stents (BMS) in coronary artery disease (CAD) (EUnetHTA 2007c), and 2) the model for diagnostic technologies (EUnetHTA 2008a) was illustrated with a Core HTA for multi-slice computed tomography for coronary angiography (CT) (EUnetHTA 2008b). Key instructions to build applications of the HTA Core Model were compiled into a handbook. Additional publications in the International Journal of Technology Assessment in Health Care were published in 2009 (Lampe, et al., 2009, Kristensen, et al., 2009).

The HTA Core Model contains nine domains originally defined in the EURASSESS project: 1) Health problem and current use of the technology (implementation level), 2) Description and technical characteristics of technology, 3) Safety, 4) Clinical effectiveness, 5) Costs, economic evaluation, 6) Ethical analysis, 7) Organizational aspects, 8) Social aspects, and 9) Legal aspects.

Each of these domains contains several topics and each topic can also have specific issues for research questions. The combination of domain, topic, and issue defines an assessment element, the basic unit of the Model. The elements are divided into *core* and *noncore* elements based on their importance and transferability (Pasternack, et al., 2009).

Drug Eluting Stents (DES) are medical devices incorporating a medicinal substance (MEDDEV Class III). The Core HTA Model on DES compared the intervention with bare metal stents (BMS) in three of

these domains: Description and technical characteristics, Effectiveness, and Costs. All the other domains framed the research questions more broadly without comparing DES and BMS.

In the Clinical *Effectiveness* domain different sources of evidence comparing DES and BMS were reviewed: HTA-reports, systematic reviews, meta-analysis and RCTs studies. The model focused on results on relative efficacy or effectiveness regarding four health outcomes: mortality (primary health outcome), morbidity, functionality/quality of life, and patients' satisfaction. Each of them was analysed in a sub-set of issues or research questions. Results on the primary health outcome (mortality) were considered completely transferable (level 3). The other three outcomes (morbidity, functionality/quality of life, and patients' satisfaction) were completely or partially (level 2) transferable across settings. Therefore, there is an underlying assumption that relative risk is the same across different settings and countries even if the baseline risk is different (EUnetHTA 2007c, p.16). In the HTA Adaptation Toolkit (EUnetHTA 2007a, p. 23), *generalisability* refers to whether the results of an HTA report can be extrapolated to other settings and *generalisable* information/data can be readily adopted. This is sometimes referred to as 'external validity'. On the other hand, *transferability* is referred as the ability to apply information and/or data from one report into a report for the user's target setting. Transferability is dependent on context specificity. This is an important distinction, which may require more discussion.

In the *Cost and Economic evaluation* domain a model to obtain cost-effectiveness estimates of DES versus BMS was performed. On the cost side, the inputs of the model were Finnish data on unit costs and resource use and the perspective employed was the one from the hospital. Therefore, only direct medical costs were considered. The outcome of the model was Quality-adjusted Life Years (QALYs) measured up to two years after the date of procedure. Neither the inputs nor the outcomes of the model were considered to be transferable across settings. QALYs may not be transferable as the values of Health Related Quality of Life (HRQoL) used to compute them may 'depend on the hospital environment, experience of the surgical team, and also selection of patients' (EUnetHTA 2007c, p. 71). On the contrary, the health outcomes reviewed in the clinical effectiveness section were considered completely or partially transferable. Despite the non-transferability of the cost-effectiveness estimates, the model was flexible enough to change the input data and create country specific estimates. The estimated ICER of the DES versus BMS obtained with Finnish data was about €100,000 per QALY. On the other hand, the results of the literature reviewed showed no consensus about the cost-effectiveness of DES vs. BMS.

In order to evaluate the potential harmonisation of HTAs requirements endorsed by the EUnetHTA Project, Trueman, et al. (2009) used the same case study comparing four different HTAs reports on drug eluting stents (DES) vs. bare metal stents (BMS) done after 2006 by three European countries (England, Austria and Belgium) and Canada. The main objective of the study was to determine the degree to which the methods adopted, evidence considered, and resulting recommendations diverged and the implications for potential harmonization of criteria across different HTA bodies. The authors observed considerable differences regarding the clinical evidence considered, the patient populations considered and the use of different sources of health outcomes data in the four reports. Despite all HTA bodies used a core dataset of Randomised Controlled Trials (RCTs) data from seven studies, they also put special emphasis in obtaining and generating evidence on clinical data and cost-effectiveness estimates at the local level using national registers and unpublished data

(England and Belgium), and field evaluation (Canada). This suggested a need for country-specific estimates before HTAs bodies did any recommendation. Despite all these detected differences, the final recommendations were fairly similar; DES was economically justified only for high-risk individuals. However, the authors identified differences in the definition of high-risk individuals. Additionally, countries may also have different cost-effectiveness thresholds.

The authors concluded there was a limited convergence in the approaches to HTA across different countries, adopting a pessimistic view of the potential for harmonisation of evidence requirements. On the other hand, they also reckoned that the HTA approach in the analysis of this specific medical device may not be representative of HTA of pharmaceuticals where there could be more room for harmonisation. In their own words 'It could be argued that the effectiveness of medical devices is more susceptible to differences in local practice patterns than pharmaceuticals. If it is accepted that pharmaceutical trials are more generalizable across settings, then there may be less need for reliance on primary evidence generation and local registry data. As such, a greater degree of harmonisation may be achievable for these technologies. Further comparisons of this type are warranted to explore whether this is the case. The current version of the HTA Core Model requires further development to guarantee its everyday usage by national HTAs bodies.'

This paper highlights some of the tensions between relative efficacy and relative effectiveness, but it is not clear how much is due to:

1. Differences as to how efficacy is translated into effectiveness (e.g. the use of QALYs);
2. Real differences in effectiveness due to local circumstances;
3. Differences in the interpretation of effectiveness data.

It is also worth noting that differences that do not change the decision may not be important.

3.2.3 The EUnetHTA Joint Action 2010-2012

In order to ensure the continuity of the EUnetHTA work beyond 2008, the organisation 'EUnetHTA Collaboration' was founded by a group of 25 organisations in 13 EU Members States, plus Norway and Switzerland as founding partners. A new grant by the European Commission ensures future work of this network that will materialise in the Joint Action 2010-2012.

The Joint Action 2010-2012 aims to maintain the collaboration and information sharing among the 25 EU member states initiated through the EUnetHTA Project. The Joint Action relies on three main objectives or streams: 1) construction of a detailed business model for collaboration addressing the sustainability of the HTA collaboration in Europe; 2) methodological developments to heighten the efficiency and transparency of HTA process in Europe; 3) applying those tools in European trans-national collaboration and at the national, regional and/or local levels. Future work on relative effectiveness of pharmaceuticals focuses important efforts in this new Joint Action.

The Joint Action is based upon the work done by EUnetHTA Project. It is also structured in eight Work Packages and most of them will continue their previous work: WP1 Communication, WP2 Dissemination, WP3 Evaluation, WP4 HTA Core Model and WP7 New Technologies. New named WPs

are: WP5 Relative Effectiveness Assessment (REA) of Pharmaceuticals, WP6 Information Management Systems and WP8 Strategy and Business Model Development.

The WP4 HTA Core Model in the Joint Action aims at creating an easy-to-use online tool and service for producing and utilizing Core HTAs. Some of the main objectives WP4 are: i) to develop a new application for screening technologies, ii) to incorporate any input from the WP5 on relative effectiveness, iii) to adapt information contained in Core HTA into local settings, using the Adaptation Toolkit developed during the EUnetHTA Project 2006-2008.

- **Work Package 5 on Relative Effectiveness Assessment of Pharmaceuticals**

The Work Package 5 on Relative Effectiveness of Pharmaceuticals is led by the College voor Zorgverzekeringen (CVZ) from The Netherlands and co-lead by the Haute Autorité de Sante (HAS) from France. The WP5 aims to develop principles and methodological guidance to improve the assessment of relative effectiveness of new pharmaceuticals. It is noteworthy that it is relative effectiveness rather than relative efficacy that is being pursued.

To that aim, the WP5 work during its three-year lifetime is structured in four main parts (EUnetHTA JA 2010). The first task will be a background review of the current processes and methodologies employed by the organisations in the Member States that are nationally responsible for the assessment of the relative effectiveness and those used in other agencies from the United States, Australia and Canada. Some of the issues to be reviewed are: i) whether the relative effectiveness assessment is developed by the HTA agency and the purpose of this assessment (clinical decision-making, reimbursement and/or for pricing decision making), ii) existence of published guidelines on relative effectiveness, iii) sources of information used in the assessment (submitted by the manufacturer or independently prepared, or other sources), iv) separation between assessment and appraisal (whether the technical assessment of relative effectiveness and the final advice for reimbursement are clearly separated or integrated in one process).

The second and the third tasks consist in developing two adapted versions of the EUnetHTA Core Model (developed by the Work Package 4) to assess relative effectiveness of new pharmaceuticals. A *Rapid version of the model* will assess a new technology at the time of introduction to the market and comparing the new technology to standard care. The *Full model* will assess (all) available technology (ies) for a particular step in a treatment pathway for a specific condition. The WP5 working plan states three key points for the development of these models: 1) be as close as possible to the national and international guidelines; 2) be non-specific context; 3) follow the principles of the HTA Core Model, although some areas may differ. Pilot assessments with a number of pharmaceuticals will test the rapid and the full models. The last main task is to produce guidelines of methodological issues for relative effectiveness assessment grouped in three main topics:

1. Comparator and comparisons: criteria for choice of the best comparator, methods of comparison, direct and indirect comparisons.
2. Outcomes: patient relevant outcomes, clinical outcomes, surrogate markers, quality of life outcomes, safety and composite endpoints.

3. Level of evidence: internal and external validity, extrapolation from efficacy results to effectiveness.

EFPIA has nominated representatives to participate in the development of both WP4 and WP5. However, it seems the industry's ability to influence the speed or content of these work packages will be very limited⁴. Limiting the opportunity for the industry to engage in EUnetHTA's processes is unlikely to help identify achievable and sustainable pan-European approaches to health technology assessment.

3.3 The Swedish EU Presidency Initiative on Relative Effectiveness

During the EU Presidency held by Sweden during 2009, the Swedish Ministry of Health and Social Affairs took an active role towards the European cooperation on assessment of the effectiveness of medical products after approval. Experts and stakeholders were brought together in the conference 'Assessing Drug Effectiveness – Common Opportunities and Challenges for Europe' to discuss opportunities and challenges of systematic cooperation across Europe on the collection and sharing of data on the effectiveness of drugs.

Examples of European cooperation in treatment areas using biologic agents were presented in chronic inflammatory diseases, cancer, and orphan diseases. When few patients in a country are exposed to a specific drug, pan-European cooperation may be important to gain knowledge on safety, efficacy and effectiveness. The general issue however is broader than patient availability. An important issue for the industry will be to understand the potential for the results of a post-launch relative effectiveness study done in one European country to be relevant to another. Having achieved acceptance (via the EMA) of the relevance of pre-launch efficacy evidence collected in one country to the rest of Europe, it is not in the interests of companies to find that they are expected to conduct post-launch studies in 27 different Member States because there is no acceptance that the results of a relative effectiveness study in one country may be applicable to patients in another country. One way of understanding similarities and differences is to conduct studies using registries in several different EU Member States and comparing the results. This constitutes the next step of the Swedish initiative after the conference. One practical example of cross-border collaboration given at the meeting was the case of tumour necrosis factor (TNF) drugs; we set this out below.

In order to provide continuation on the Swedish Initiative, the Swedish Medical Products Agency (MPA) took the lead during 2010 on behalf of the Swedish Ministry of Health and Social Affairs. It is expected that this initiative will be continued during the next two-three years by future EU presidencies.

An Oversight Committee (OC) has been set up to supervise work on a number of pilot projects which aim to collect data and disseminate evidence on relative effectiveness across different interventions and disease areas in the EU. The OC gathers a wide range of stakeholders such as EFPIA, EMA,

⁴ Personal communication.

EUnetHTA, NICE and the permanent representation of Sweden to the EU among others. It is expected that patient groups and clinical experts will also join the OC as key stakeholders.

The OC selected four pilots in April 2010: i) EUROFEVER, a registry collecting outcome data of patients affected by major autoinflammatory diseases in childhood; ii) EUTOS registry, which is part of the project EUTOS for Chronic Myeloid Leukaemia (CML); iii) establishment of a new Multiple Sclerosis Treatment registry by the European Federation of Neurological Societies (EFNS); and iv) European Biologics Registers for Rheumatic Diseases, a collaboration of three registers to collect outcome data on biologic agents for Rheumatoid Arthritis and explained in more detail below.

The role of the OC is to support the collection and data sharing within each network, the analyses of treatment outcomes are matters for the respective networks. The OC work plan with each pilot project will include conducting surveys, compilation of primary data collection, consultation with experts, advice on methodological or regulatory matters, and contributions to scientific standards.

After the pilot projects are finished, the intention is that the OC will continue to contribute to the improvement of the content and data management of the networks. The aim of this work is to gain knowledge on how drug effectiveness data can be collected and shared in a European setting (EFPIA, 2010).

European biologics registers for rheumatic diseases

At the Swedish Initiative conference Professor Symmons (University of Manchester) presented work on three European Biologics Registers for Rheumatic Diseases that monitor the long-term safety of the first tumour necrosis factor (TNF) α inhibitors. These three registers, from the UK, Germany and Sweden⁵, have agreed on a standardised reporting system for serious adverse events. They follow patients irrespective of treatment continuation with a specific drug and the majority of these patients are expected to be exposed to more than one drug in future given the increasing number of available biological drugs (Zick, et al., 2009).

These three registers are part of the European League against Rheumatism (EULAR), a platform representing the patient, health professional and scientific societies of rheumatology of all the European nations. Under its auspices, biological registers of European countries have been working closely: UK, Sweden, Germany, Spain, Norway, Denmark, The Netherlands and Switzerland.

This international collaboration is crucial to gaining substantial knowledge on very rare events as national registers may not have enough of each of these cases. Also, it allows for the necessary harmonisation of methodologies to compare results across the countries. Observational studies pose important challenges in their statistical analysis. In this context, confounding by indication can produce biased results as prognostic factors may influence treatment decisions, hence unbalancing the baseline risk of the treatment and control groups. A further difficulty arises from the multidrug exposure of the patients in the registers as the effectiveness of one drug cannot be easily isolated from the treatment effect of other drugs consumed by the same patient.

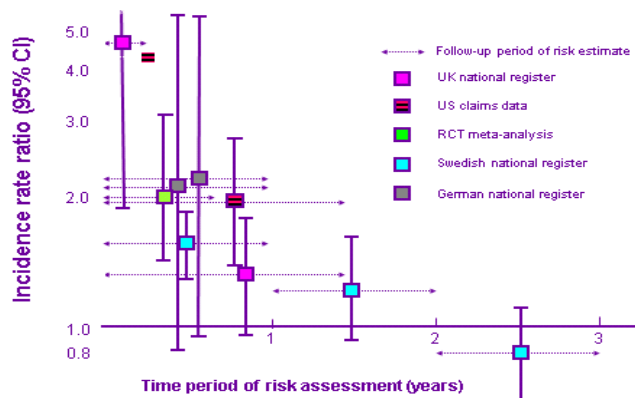
⁵ The British Society for Rheumatology Biologics Register (BSRBR), the German Biologics Register (RABBIT), and the Swedish Biologics Register (ARTIS).

Results from these three registers have been reported elsewhere (Askling and Dixon, 2008). The authors reviewed studies on the safety of anti-tumour necrosis factor (TNF) drugs in patients with rheumatoid arthritis (RA) in five different areas: infection, malignancy, ischemic heart disease, interstitial lung disease, and death. Regarding infections, results from observational studies using data from the British (BSRBR), German (RABBIT) and Swedish (ARTIS) registers were compared with results from a meta-analysis using RCTs data and two observational studies from the United States.

The studies differed in the length of follow-up, the study populations, the definition of ‘serious infection’, and the calculation of ‘time at risk’. The authors controlled for these differences and plotted the relative risk estimates against treatment duration (Figure 3.1), showing a consistent figure of higher risks estimates during the first months of treatment followed by less elevated or reduced risks estimates with increasing duration of the treatment.

However, Askling and Dixon (2008) reckoned the interpretation of the pattern was not straightforward. This early increased risk could be due to biological factors or different sources of bias (for example, a lower threshold for treating infections early in therapy, or patient drop-out). The authors also pointed out the differences in comparison groups across studies and its impact on markedly different estimates of relative risk despite similar rates of serious infection in the two anti-TNF treated cohorts (for example, the German study estimated a doubled risk, while the UK study did not detect any increase in risk).

Figure 3.1. Relative risk of serious infection in patients with rheumatoid arthritis treated with tumour necrosis factor antagonists, by time since start of treatment



Source: Askling and Dixon, 2008. This version of the graph was presented by Symmons (2009).

Although the registers only collect data on the safety of TNF drugs, this collaborative work represents a good example of commitment of various European countries to standardise methods and compare results from observational data.

3.4 Next steps

It is clearly essential that the industry be engaged in the various pan-European initiatives that are taking place. However, there appears to be a lack of clarity as to the issues that EUnetHTA and others expect to be raised and what the results of the exercise might be.

3.5 References

Asking, J. and Dixon, W. (2009) The safety of anti-tumour necrosis factor therapy in rheumatoid arthritis. *Current Opinion in Rheumatology*. 20(2), 138–144.

European Federation of Pharmaceutical Industries and Associations (EFPIA). (2010) Background note on the 'Swedish Initiative'. Brussels: European Federation of Pharmaceutical Industries and Associations.

EUnetHTA. (2007a) *Work Package 5. EUnetHTA HTA Adaptation Terms v. 2 and Glossary of HTA Adaptation Terms*. Available at http://www.eunethta.net/upload/Barcelona_2007/AdaptationToolkit%20%20Glossary%20combinedJune2007.pdf [Accessed 10 August 2010].

EUnetHTA. (2007b) *Work Package 4. HTA Core Model for medical and surgical interventions. First Public Draft*. Available at http://www.eunethta.net/upload/WP4/EUnetHTA_WP4_CoreModelforInterventions_FirstPublicDraftRevised-2007-07-11.pdf [Accessed 10 August 2010].

EUnetHTA. (2007c) *Work Package 5. Core HTA on Drug Eluting Stents (DES). First Public Draft*. Available at http://www.eunethta.net/upload/WP4/EUnetHTA_WP4_CoreHTAonDES_FirstPublicDraftRevised-2007-07-11.pdf [Accessed: 10 August 2010].

EUnetHTA. (2008a) *Work Package 4. HTA Core Model for diagnostic technologies. First Public Draft*. Available at http://www.eunethta.net/upload/WP4/Validation/EUnetHTA%20WP4%20-%20Core%20Model%20for%20Diagnostic%20Technologies%20-%20DRAFT%20-%202008-07-15_2.pdf [Accessed 10 August 2010].

EUnetHTA. (2008b) *Work Package 4. Core HTA on multislice computed tomography (MSCT) coronary angiography. First Public Draft*. Available at <http://www.eunethta.net/upload/WP4/Validation/EUnetHTA%20WP4%20-%20CoreHTAonMSCT%20-%20DRAFT%20-%202008-10-07.pdf> [Accessed 10 August 2010].

EUnetHTA. (2009a) *EUnetHTA Project. Overview of results. Years 2006-2008*. Available at http://www.eunethta.net/Public/Work_Packages/EUnetHTA-Project-2006-08/EUnetHTA_Deliverables_project_2006-2008/ [Accessed 10 August 2010].

EUnetHTA. (2009b) *European network for HTA Joint Action. Technical Annex Part 1*. Available at: http://www.eunethta.net/upload/Joint%20Action/JA%20WP1/GrantAgreement%20Tech%20Annex/JA_Annex%201%20-%20part%201.pdf [Accessed 10 August 2010].

EUnetHTA. (2010) *Work Package 5. Relative Effectiveness Assessment (REA) of Pharmaceuticals Fact Sheet*. Available at http://www.eunetha.net/upload/Joint%20Action/WP5%20REA%20of%20Pharmaceuticals/Factsheet%20WP5_Aug2010.pdf [Accessed 10 August 2010]

EUnetHTA Joint Action. (2010) *EUnetHTA Joint Action on HTA 2010-2012. 3-year Work Plan. May 7, 2010*.

H LPG. (2008) *High Level Pharmaceutical Group 2005-2008. Conclusions and Recommendations*.

Kristensen, F.O., Lampe, K., Chase, D.L., Lee-Robin, S.H., Wild, C., Moharra, M., Garrido, M.V., Nielsen, C.P., Røttingen, J.A., Neikter, S.A. and Bistrup, M.L. (EUnetHTA). (2009) Practical tools and methods for health technology assessment in Europe: Structures, methodologies, and tools developed by the European network for Health Technology Assessment, EUnetHTA. *International Journal of Technology Assessment in Health Care*. 25(Suppl 2), 1-8.

Lampe, K., Mäkela, M., Velasco Garrido, M., Anttila, H., Autti-Rämö, I., Hicks, N.J., Hofmann, B., Koivisto, J., Kunz, R., Kärki, P., Malmivaara, A., Meiesaar, K., Reiman-Möttönen, P., Norderhaug, I., Pasternack, I., Ruano-Ravina, A., Räsänen, P., Saalasti-Koskinen, U., Saarni, S.I., Walin, L. and Kristensen, F.B. (EUnetHTA). (2009) The HTA Core Model: A novel method for producing and reporting health technology assessments. *International Journal of Technology Assessment in Health Care*, 25(Suppl 2), pp. 9-20.

Pasternack, I., Anttila, H., Mäkelä, M., Ikonen, T., Räsänen, P., Lampe, K. et al. 2009. Testing the HTA Core Model: Experiences from two pilot projects. *International Journal of Technology Assessment in Health Care*. 25(Suppl 2), 21-27.

Swedish EU Presidency. (2009) *Assessing Drug Effectiveness – Common Opportunities and Challenges for Europe*. Post Conference Booklet. 29 July. Available at http://www.lakemedelsverket.se/upload/eng-mpa-se/drug-effectiveness/Drug_Effectiveness_2009-Post-Conference-Booklet.pdf [Accessed 10 August 2010].

Trueman, P., Hurry, M., Bending, M. and Hutton, J. (2009) The feasibility of harmonizing health technology assessments across jurisdictions: A case study of drug eluting stents. *International Journal of Technology Assessment in Health Care*. 25(4), 1-8.

Zink, A., Askling, J., Dixon, W. G., Klareskog, L., Silman, A.J. and Symmons, D.P.M. (2009) European biological registers: Methodology, selected results and perspectives. *Annals of Rheumatic Diseases*. 68, 1240-1246.

4 Systematic Literature Review on Relative Effectiveness in Europe

4.1 Introduction

The primary objective of this literature review was to find us understand the extent of likely variation if any in underlying relative efficacy and relative effectiveness of drugs used in one or more of the 27 Member States. We therefore looked for evidence of heterogeneity in relative effectiveness of any drug and its explanatory factors across the 27 European Union (EU) Member States. Appendix 2 shows the approach used for the literature review. Following the definition of the EU-High Level Pharmaceutical Forum (HLPF), as stated at the beginning of the report, *relative effectiveness* is the extent to which an intervention does more good than harm compared to one or more intervention alternatives for achieving the desired results when provided under the *usual circumstances* of health care practice. This is usually measured in observational and post-launch studies. In the absence of evidence on relative effectiveness, the review also considered studies analysing *relative efficacy*, which refers to the extent to which an intervention does more good than harm, under *ideal circumstances*, compared to one or more alternative interventions. This is measured in experimental settings as RCTs. Even if the difference between both effectiveness and efficacy is clear in theory, the terminology is used inconsistently in the literature. In some cases, authors just refer to the treatment effect.

The literature review searched for studies analysing variations in relative effectiveness or efficacy and questioning the appropriateness of generalising results across different countries when such differences are encountered. The term *generalisability* has different meanings depending on the context where it is employed. In clinical evaluation literature, it mainly refers to the characteristics of patients in a given study and how representative they are of a broader population with different characteristics. In comparison, in economic evaluations, generalisability is usually defined as the extent to which the results of a study based on measurement in a particular patient population and/or specific context hold true for another population and/or different context (Sculpher, et al., 2004).

Generalisability and *transferability* are terms usually used interchangeably in the literature. To avoid confusion, the EUnetHTA Project⁶ definitions of both expressions, which we referred to earlier, are employed in this review. 'Generalisability' refers to the extent to which the overall results of an HTA report can be readily adopted by other settings while 'transferability', a subset of generalisability, is the ability to apply particular information and/or data from one report to another report for the user's target setting. Transferability is dependent on context specificity. The more context specific, the less likely that data/information in one report can be adopted by another, i.e. transferred without making any changes or additions.

⁶ EUnetHTA, (2007), Work Package 5. EUnetHTA HTA Adaptation Terms v. 2 and Glossary of HTA Adaptation Terms. http://www.eunetha.net/upload/Barcelona_2007/AdaptationToolkit%20%20Glossary%20combinedJune2007.pdf [Accessed 10 August 2010].

Transferability of economic data across jurisdictions from cost-effectiveness studies has been debated extensively in the literature. A recent publication by the ISPOR Good Research Practices Task Force (Drummond, et al., 2009) analysed and discussed important methodological and practical issues in the field. Recommendations for good research practices were provided, including analysis of individual patient data (using regression analysis or multilevel models) and decision-analytic modelling both to establish the extent of transferability and to use the data in the best way on another jurisdiction.

The generalisability of relative effectiveness results of a specific drug across different countries could be compromised by different factors including patients' characteristics, the comparators against which the drug is evaluated, the health outcome measures used, the valuations given to health outcomes, variations in clinical practice, and resource allocation. However, generalisability of relative efficacy results across different locations is broadly assumed in the medical literature. The reason is that strict protocols in RCTs 'in principle' control for all such sources of potential variability. Despite this, methodologies in RCT are not exempt from difficulties and selection bias can be an obstacle for generalisation of results as most patients are selected for randomisation (either by themselves or by clinicians) rather than randomly drawn. There may be unobservable differences between the trial sample and the wider population generating a biased estimate of the average treatment effect (Manning and Claxton, 1997).

Previous studies have reviewed the heterogeneity of treatment effects and generalisability of trial results across different patient subgroups. O'Connell and colleagues (2001) conducted a literature review on the applicability of results of RCTs and systematic reviews of RCTs. The main objectives were to identify the current methods used to apply clinical trials results to usual practice, assess the quality of these methods, and propose appropriate techniques for applying clinical trials results to individual patients based on a consideration of benefit and risk.

One of the issues addressed the O'Connell study was the common underlying assumption that the relative treatment effect remains constant across subgroups receiving the intervention as biological factors should not vary that much among humans. This assumption is correct in so far as there is no prior justification and no strong evidence confronting this premise. The authors emphasize that the evidence should come from testing whether a specific factor modifies the treatment effect, and not by testing within each individual subgroup as has been commonly done in the literature (which can produce both false-positive and false-negative results). This ideally would be done with individual patient data within the same trial. Otherwise, variation could occur as the result of differences in trial design when comparing data between trials.

Potential sources of heterogeneity in treatment effect were also theoretically identified in that study and are featured here in Box 2.

Box 2. Sources of underlying variation in treatment effects:

- a) patient characteristics (e.g. age, gender, biochemical markers)
- b) intervention features (e.g. the timing, compliance, or intensity)
- c) disease features (e.g. the severity, epidemiology) and
- d) the measure of the effect used (relative risk versus risk difference) and clinical endpoints.

When comparing several studies evaluating a specific intervention, other heterogeneity also may occur due to differences in trial design (setting, subjects, comparators, length of follow-up, outcomes measured, etc.).

The authors reckoned that tests for heterogeneity of treatment effect are usually underpowered and thus heterogeneity is not detected even if present. Therefore, tests to explore statistical significant relationships between each of the potential effect modifiers (factors causing variation) and the relative treatment effect should be performed regardless of the initial result of the heterogeneity test.

The authors also drew attention to another important assumption common in the literature that should be also checked: whether the relative measures of the treatment effect (such as relative risk reduction) vary with baseline risk level, i.e. if it is lower (or may be higher) in patients with lower baseline risk. Box 3 provides definitions of these measures. The authors found that the relative risk reduction appeared to be constant in many interventions, but not all (e.g. patients with carotid artery stenosis treated with carotid endarterectomy, and class I anti-arrhythmic drugs).

Box 3. Absolute and relative measures of treatment effects.

The absolute risk reduction (ARR) or the absolute treatment effect is the number of events prevented. Two equivalent specifications:

As the absolute difference in outcomes rates between the control (CER) and the treatment groups (EER): $ARR = CER - EER$

As the interaction of i) the baseline or reference risk in the control group, that is the population's exposure to a specific event (CER assuming the control group is representative of the population), and ii) the relative risk reduction (RRR), that refers to the reduced exposure to the event due to the drug or treatment effect.

$$ARR = CER * RRR \text{ where } RRR = (CER - EER) / CER$$

Two examples of ARR measures are the differential mortality and survival rates.

The authors identified some areas for further research. Firstly, in order to explore and discuss variations in relative treatment effects due to effect modifiers or differences in baseline risk, large pragmatic trials with heterogeneous patient populations are needed to facilitate this analysis. Also, those researchers who are 'undertaking systematic reviews should examine heterogeneity, effect

modification and the importance of baseline risk rather than concentrate on producing pooled (average) estimate of treatment effect'. Finally, further methodological research is required to produce more efficient statistical tools. Not discussed, however, is whose responsibility is it to conduct this research. We believe that it is the responsibility of all stakeholders involved, including regulators, HTA agencies and industry.

O'Connell, et al. did not focus on the generalisability of effectiveness results across different countries. The objective of our review was to identify literature studying variations in relative effectiveness and efficacy of drugs in any therapeutic area in two or more EU countries, and, if heterogeneity was detected, identify its underlying explanatory factors.

The results show little available evidence in this field. However, as this is a topic of increasing importance on the European Commission agenda, several initiatives are underway. The HAPPY AUDIT (Bjerrum, et al., 2010) is a multinational study involving general practitioners from six countries (Denmark, Sweden, Lithuania, Russia, Spain and Argentina) aiming to evaluate the impact of a multifaceted intervention to reduce the inappropriate use of antibiotics in primary health care. The project is ongoing and no results have yet been reported. However, relatively new statistical techniques, such as multilevel analysis, will be employed to test whether environmental and organisational factors influence the effectiveness of the intervention across the participating countries. (Questions to be answered include: Are there any significant differences among regions? Are there some characteristics of the GPs influencing the intervention results?) This study, however, was not included in our results as it is a public health intervention.

4.2 Methods

A systematic literature review of medical and health economics literature was conducted using a combination of database and bibliographic searchers. Four main databases were searched for the period January 2000 to May 2010: Medline, Embase, EconLit, and Health Management Information Consortium (HMIC). A search in Google Scholar also was conducted. Our search covered only English language studies and logical combinations of keywords related to effectiveness, generalisability, external validity, transferability, Europe and review were searched in titles and abstracts.

Papers identified as being potentially relevant from titles and/or abstracts were downloaded from the four databases into a Reference Manager database; some citations were introduced manually.

The papers included could be either methodological, empirical studies or reviews analysing heterogeneity of effectiveness or efficacy of any drug in two or more EU countries. Excluded papers were those analysing health interventions and technologies other than drugs. Studies comparing pooled results for Europe with other regions, such as the United States or Australia, also were also excluded as the aim was to explore differences between the EU Member States.

A total of 326 articles were obtained after eliminating duplicates and titles and abstracts were screened. Forty-four studies were judged to be of potential relevance and, on further assessment, ten studies formed the bases of the review and annotated bibliographies. Methodological and empirical analyses, as well as reviews, were considered.

4.3 Results

The literature on heterogeneity of relative effectiveness and efficacy in Europe was scarce and difficult to identify. There are several possible reasons for this. Firstly, the underlying assumption is that clinical outcomes in European trials are likely to be generalisable across the continent (Willke, et al., 1998) and few studies to date have questioned this premise. Secondly, attempts to test this assumption encounter statistical difficulties. Multinational clinical trials are powered to determine the overall treatment comparison and they are underpowered to detect differences in subgroups (O'Shea and Califf, 2001). Sample sizes in each of the sites usually are too small to detect statistically significant differences in relative efficacy even if they exist. This lack of statistical power is due partly to the fact that multinational trials do not explicitly assess differences across sites as well as the difficulty in recruiting patients and the high costs involved in these studies. Tests comparing subgroups have proved to be misleading and new statistical techniques are needed to make these comparisons (O'Connell, et al., 2001). Multilevel models, one of the new developments in this area, are discussed later in this section.

The search did not identify observational studies meeting our inclusion criteria. The development of this type of study may be difficult for two main reasons: a) registry data is costly and complicated to collect and, when available, it usually is not easily accessible and b) databases do not record patient outcome and resource allocation (Jönsson, 2009). Further collaboration among the Member States in sharing patients' registries will be crucial to produce more evidence in this field in future.

Observational studies to estimate drug relative effectiveness also face methodological problems and harmonisation of criteria is essential to making any comparison (between and within countries). Differences in health outcome measures (clinical endpoints, markers, etc.), patient characteristics, and comparators, among other issues, pose serious difficulties for any attempt at comparison. Selection bias in these studies is also a very common issue. For example, in the case of the European Biologics Registers for Rheumatic Diseases (Zick, et al., 2009), confounding by indication can produce biased results as prognostic factors may influence treatment decisions leading to different patient characteristics in the treatment and the control groups. Additionally, drop-outs by patients with poor prognosis may lead to a better average prognosis for those patients remaining on therapy, influencing the results on treatment effectiveness. A further difficulty arises from the multidrug exposure of patients in the registers as the effectiveness of one drug cannot be easily isolated from the treatment effect of other drugs consumed by the same patient.

Table 4.1 summarises the main features and findings of the ten relevant studies included in the review. The search identified one study (Smith, et al., 2008) strictly analysing differences in treatment effects using efficacy data in two geographical areas in Europe (Western Europe, and Central and Eastern Europe). Smith, et al. (2008) reported the results of a multinational RCT using Darbepoetin Alfa (DA) against placebo for the treatment of anaemia in patients with active cancer not receiving chemotherapy or radiotherapy. The study enrolled patients (989 in total) from 144 sites in Europe (75% of total patients), North America and Australia (referred as 'rest of world', ROW, in the analysis) from 2004 to 2006. The primary clinical endpoint was all occurrences of transfusions from weeks five through 17 and survival was one of the safety end points of the study (during the

pre-treatment phase and a two-year follow up period). Regional differences were reported in the survival analysis showing that patients from Western Europe had worse survival (ROW). However these differences were not statistically significant (at the 95% confidence interval).

To overcome the sparseness of the literature on heterogeneity of relative effectiveness across European countries, we extended our scope to include studies analysing variations in cost-effectiveness to explore whether these studies also provide information on differences in relative effectiveness. Because of the potential interaction of clinical and cost data, such studies could constitute a valuable source of information about variations in effectiveness and its potential sources of heterogeneity.

Table 4.1. Summary of studies included in the literature review.

Study	Type	Disease area	Intervention	Setting	Primary clinical endpoint	Relevant findings
Smith, et al., 2008	RCT 2004-06	Anaemia in patients with active cancer (not receiving chemotherapy/radiotherapy)	Darbepoetin Alfa (DA) vs placebo	Western Europe (WE), Central & Eastern Europe, ROW (North America and Australia)	All occurrences of transfusions from week 5 through 17	WE patients treated with DA had worse survival outcomes (safety endpoint) but these differences were not statistically significant No analysis of factors underlying these differences
Sculpher, et al., 2008	Review of CE studies	Non-specific	Non-specific	Non-specific	Different depending on the disease areas	Most cited factors underlying variation in CE results: unit costs, variations in clinical practice, geographical setting, and healthcare resources
Barbieri, et al., 2005	Review of CE studies	Non-specific	Different pharmaceutical treatments	UK, Spain, Germany, France, and Italy	Different depending on the disease areas	Three studies reported heterogeneity on relative effectiveness but no sources of this variation were further explored Significant but not systematic differences in CE results due to unit costs and resource use
Hakkart-van Roijen, 1998	CE analysis	Chronic plaque psoriasis	Tapered vs abrupt discontinuation of cyclosporin	UK, Spain, Turkey, and Canada	Total days of systematic therapy-free days (STDFs)	Not statistically significant differences in STDFs across the four countries because the small number of patients in each of them
Wilke, et al., 1998	CE analysis	Aneurysmal subarachnoid haemorrhage (SAH)	Four arms: three different doses of tirilazad + a vehicle-only arm	Five countries from a group of nine European countries, NZ, and Australia	Mortality rate	Factors underlying differences in mortality rates across countries: severity and patient characteristics

Study	Type	Disease area	Intervention	Setting	Primary clinical endpoint	Relevant findings
Cook, et al., 2003	CE analysis from the 4S trial	Cholesterol	Simvastatin vs placebo	Denmark, Finland, Iceland, Norway, and Sweden	Mortality rate	Mortality rate across countries was not statistically different to the overall mortality rate Therefore health outcomes could be pooled
Pinto, et al., 2005	CE analysis from the ASSENT-3 Trial	Cardiovascular	Three new thrombolytic regimes: Heparin, Enoxaparin and Abciximab	26 countries (15 are EU Member States)	Freedom from death, in-hospital reinfarction and refractory ischaemia for 30 days	Detected between-country heterogeneity on effectiveness results but countries these were not identified In this situation, the proposed method provides more efficient estimates than pooled estimators reducing the standard error
Manca, et al., 2007	CE analysis from the ATLAS trial	Chronic heart failure	Low vs high dose of ACE inhibitor lisinopril	17 countries from a group of 16 European countries, US, Canada, and Australia	Survival gain	Both patient and country-level factors explained variability in differential survival and costs across countries Country-level factors: life expectancy, private and public health expenditure, alcohol and tobacco Patient factors: age, sex and left ventricular function
Manca and Willan, 2006	Methodological paper CE analysis	Non-specific Some case studies in cardiovascular interventions are provided	Non-specific Case studies from the ATLAS trial and WOSCOPS study	Non-specific	Non-specific	The authors proposed an algorithm to assist the choice of appropriate analytical strategy to adapt CE results from different countries

Study	Type	Disease area	Intervention	Setting	Primary clinical endpoint	Relevant findings
Vemer, et al., 2010	CE modelling study.	Smoking-related diseases.	Four arms: three Smoking Cessation Therapies (SCT) + unaided arm.	Holland, Belgium, UK, Germany, Sweden, and France.	N/A	Between country variability in CE results mainly due to three factors: discount rates, incidence and mortality of smoking-related diseases, and utility values used to estimate QALYs. Other six analysed factors showed lower or null effect on the heterogeneity of the results: demography, smoking prevalence, all-cause mortality, costs of disease, resources used for SCTs, unit costs of SCTs.

Sculpher, et al. (2004) conducted a systematic literature review of methodological studies analysing factors affecting variability and empirical papers estimating geographical variability in economic evaluations. Table 4.2 lists the factors resulting from the review of empirical studies, which included studies using patient-level data and modelling techniques. A total of 33 references were included in the review and most usually identified more than one factor in the analysis. For instance, one of the papers⁷ identified seven factors causing variations in cost-effectiveness results. The most frequently factor identified that affected the results of economic evaluations was variation in unit costs, followed by variations in clinical practice, geographical setting and availability of healthcare resources. In addition to those in Table 4.1, the review of methodological studies added other factors such as capacity utilisation, economies of scale and incentives in the health care system.

Table 4.2. Number of studies citing specific factors causing variation in cost-effectiveness results

Factors	Number of studies
Patient/individual factors	
Epidemiology	4
Demography	5
Case mix	1
Compliance	2
Clinical Factors	
Artificial study conditions	6
Clinical practice variation	28
Timing of assessment	1
Skills/experience	3
Health Care System	
Healthcare system	4
Treatment comparators	2
Absolute/relative costs (prices)	31
Opportunity cost	1
Wider socio-economic factors	
Healthcare resources	8
Culture/attitudes	2
Geographical setting	27
Industry-related bias	1
Technological innovation	1
Perspective	1

Source: Sculpher, et al., 2004. Note: Case mix refers to clinical and socio-demographic characteristics of patients undergoing treatment.

⁷Bennett, C.L., George, S.L., Vose, J.M., Nemunaitis, J.J., Armitage, J.L., Armitage, J.O., et al., (1995) Granulocytemacrophage colony-stimulating factor as adjunct therapy in relapsed lymphoid malignancy: implications for economic analyses of phase III clinical trials, *Stem Cells*, 13, 414–20.

The cost-effectiveness studies reporting variation in clinical practice⁸ mainly refer to differences across countries instead of differences between experimental settings and standard practice. For this reason, all except two papers⁹ in this group are the same as those studies reporting differences across geographical settings. The cost-effectiveness studies reporting variations across countries were explored further in the present review as these variations can potentially affect the efficacy or effectiveness of drugs. Contrary to this idea, it is argued that these differences are more likely to have an impact on medical interventions like surgery rather than on pharmaceuticals where the process is less influenced by the clinician's experience (Sculpher, et al., 2004).

Seven of these studies were not explored further as they did not meet our inclusion criteria; some did not evaluate drug therapies while others did not analyse potential differences across EU countries (comparing Europe as a whole region with other regions or only one EU country with a non-EU country).

Among the 21 studies that remained, three used individual patient data in their analysis while the remaining 18 were modelling studies. The majority of these cost-effectiveness studies looked at the hospitalised, etc.) on resource use and consequently on treatment costs in each of the included countries. Unfortunately, they did not analyse the effect of the differences in medical practice on the efficacy or effectiveness of the drugs. This is particularly so for modelling studies, where clinical data usually are obtained from meta-analysis or clinical trials assuming the same relative risk reduction across countries. In the reviewed modelling studies, the authors accounted for differences in health resource consumption to obtain more accurate country-specific costs estimates.

Some of the studies did not report detailed analysis of variation in clinical practice, but only mentioned this as a potential limitation to generalising the studies' results to other settings. Among those providing more detail, they usually referred to differences across countries in length of hospitalisation, number of tests, number of visits, and primary versus secondary care.

Drummond (1994) pointed out the importance of considering the potential effects of country variations in clinical practice and resource use on the efficacy of a therapy. The author questioned the appropriateness of pooling clinical results from a wide range of settings when different resource inputs have been used. This is especially relevant in ancillary care, where protocols are often less standardised, or intensive care where professionals' experience and resources are as important as the evaluated drug. However, the author concluded that, in his model of drugs for acid-related diseases, this issue may have limited relevance as the main health outcome relies on the drug's efficacy, which is supposed to vary less across settings.

Barbieri, et al. (2005) also conducted a literature review on the causes of variation in cost-effectiveness results exclusively focusing on pharmaceutical treatments in two or more European

⁸ Reference numbers of these 28 studies in Sculpher, et al. (2004) are: 31, 56, 78–81, 83, 84, 87, 89–103, 105, 106, 115, 116.

⁹ The two papers were not included because they did not meet the inclusion criteria; one was a regional comparison of US vs Europe and the second was a multicentre trial only based in France.

countries. The results showed significant but not systematic differences in cost-effectiveness estimates among the five major European countries (UK, Germany, France, Italy and Spain); this implies that the generalisability of results from country to country was not straightforward. The authors argued, however, that the implications of these variations depend ultimately on the decision makers; a common high threshold per QALY in Western European countries would mean these differences might not be significant as many drugs would be considered cost-effective. In most of these studies (27) the same effectiveness across countries was assumed. For the remaining studies (17), only three¹⁰ (Hakkart-van Roijen, et al., 1998, and Willke, et al., 1998) reported differences in effectiveness as an important factor influencing cost-effectiveness results. Differences in unit costs and resource were found to be the key drivers of variations in cost-effectiveness.

One of three studies cited in Barbieri, et al. (2005) was an empirical study by Hakkart-van Roijen, et al. (1998). The authors assessed the costs-effectiveness of non-blind multinational clinical trial tapered-versus-abrupt discontinuation of microemulsion formulation of cyclosporin in patients with chronic plaque psoriasis in Canada, Spain, Turkey and the UK. Differences in health outcome - measured were evident in the total days of systematic therapy-free days (STFDs), but proved not be significant because the small number of patients in each of the four countries limited the statistical power of the analysis. However, in the four countries, 'tapered' was more effective than 'abrupt' discontinuation and the former was dominant in Spain, UK, and Turkey. In Canada, it produced an incremental cost-effectiveness ratio of C\$1.4 per STFD. In Spain, the difference in costs between the two arms was due to differences in severity, and consequently in efficacy, not dissipated by the randomization.

Regarding methodological studies in this area, one important contribution is by Willke, et al. (1998). They proposed a method for evaluating the generalisability of the overall cost-effectiveness results from a multinational clinical trial to individual countries participating in the trial. They used a regression analysis introducing country-specific variables to control for differences in costs and outcomes (fixed effects models). To illustrate their methodology, they took patient data from a RCT for aneurysmal subarachnoid haemorrhage (SAH) with four arms (three different doses of tirilazad plus a vehicle-only arm) run in nine different European countries, Australia and New Zealand. From those, they included five countries for which they had economic data on hospital costs. Unfortunately, these countries were not identified, but included between three and five European countries, depending on whether Australia and New Zealand were included in the analysis. Initial mean tests in outcome measure (mortality rates) showed statistically differences across the five countries. However, after controlling for severity and patient characteristics, these differences diluted therefore concluding outcome results were transferable across settings (although statistical power of the analysis was limited)¹¹. On the contrary, cost-effectiveness ratios showed persistent

¹⁰ Two of these studies, Hakkart-van Roijen, et al., 1998 and Willke, et al., 1998, could be found and included in the literature review. The third one could not be identified.

¹¹ The original publication of the trial results (Kassell, et al., 1996) did not report treatment-by-country results and outcome data was pooled for the analysis.

differences across the five countries providing evidence that adjustments were necessary before extrapolating the overall cost-effectiveness estimates to a specific country.

Cook, et al. (2003) proposed existing tests for interaction of clinical endpoints that could be used to analyse economic endpoints, cost-effectiveness and net health benefits in data from multinational trials. The authors illustrated these methods using data from the 4S trial, a randomised double-blind study comparing simvastatin, cholesterol lowering therapy, and placebo in the five Scandinavian countries: Denmark, Finland, Iceland, Norway and Sweden. The patients (4,444 in total) had a prior myocardial infarction and/or had stable angina pectoris.

The authors analysed potential variations across countries in mortality rates, hospitalization and cost-effectiveness ratios. Heterogeneity of treatment effect across countries was referred as treatment-by-country interaction and two categories were defined: i) qualitative (or cross-over) interaction, when there are differences in the effect sign/direction; and ii) quantitative (or non cross-over) for differences in the magnitude of the effect but with the same sign.

Clinical results extracted from cost-effectiveness studies need to be treated carefully as cost-effectiveness ratios use absolute differences in effectiveness and not relative measures of effectiveness. Cook, et al. (2003) emphasized on this issue: 'separate evaluations of homogeneity among countries for treatment effects in clinical outcomes and utilization variables are not adequate to ensure homogeneity in the cost-effectiveness ratios. For example, models used to evaluate treatment effects often use relative measures such as the odds ratio, or percentage reduction, whereas incremental cost-effectiveness ratios use the ratio of absolute differences. Heterogeneity in absolute treatment effects (measured as a difference) can occur when there are large country-to-country differences coupled with a constant multiplicative treatment effect (homogeneity in relative treatment effects). Other scenarios can be constructed; these are further complicated by the fact that costs and benefits are not independent and that pricing can significantly affect health care utilization.'

Statistically significant treatment effect by country interactions are generally not detected in clinical data. This may be because the biological effect does not vary across settings, as commonly believed, or because of the sample size in each centre is too small to provide statistical power to the analysis. The authors estimated the mortality rate for each of the five countries and found strong treatment and country effects (Denmark doubled the proportion of deaths recorded in Finland), but no treatment-by-country interaction (the proportion of deaths in each individual country was similar to the overall treatment impact). The results therefore suggested lack of interaction (either qualitative or quantitative) in the relative effect of the treatment in mortality, which allowed the analysts to pool the treatment effect estimates over countries to gain statistical power and precision in the analysis. Similarly applied to hospitalizations, no interactions were detected.

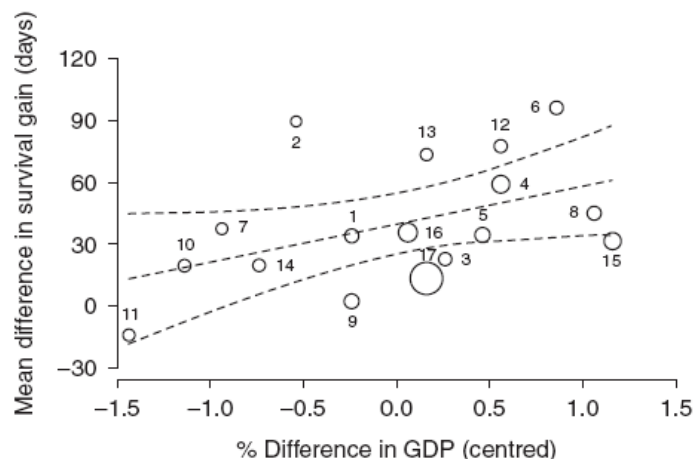
Cook, et al. found no evidence of interaction in the incremental cost-effectiveness ratio (ICER) and economic data could be pooled using different weights to obtain an overall ICER. However, the authors used Swedish prices to estimate drug and hospitalisation costs in the five countries as not completed pricing data was available. Therefore, differences in the ICERs were not driven by prices, but by differences in resource use (numerator) and absolute difference in the treatment effect (denominator).

The authors pointed out the importance of trial design in order to detect heterogeneity in both relative treatment effect and economic measures: sample size, country and centre selection, and outcome variable selection are crucial for validity and generalisability of trials' results.

Pinto, et al. (2005) reckoned the main advantages of multinational trials are to increase statistical power by using larger sample sizes and to produce a perception of greater generalisability. However, these advantages rely on the assumption of similar effectiveness across countries to pool the health outcomes results. Even if this could be achieved by strict protocols, the authors noted, it is more controversial for cost results. The authors proposed a Bayesian approach (shrinkage estimates) to estimate country-specific costs-effectiveness results more efficiently than by using the country-specific data alone. They conducted a simulation based on summary estimates of country level differential costs and effects from an RCT run in 26 countries to assess the safety and efficacy of new thrombolytic regimens (ASSENT) comparing three different interventions. The between-country variability in differential effects was even greater than the corresponding variability in differential cost. The authors concluded that pooled estimators were not appropriate of country-specific effects in the cases where country heterogeneity was detected. Also, their estimators reduced the average standard error between 20% and 59%.

Recent developments in methods have been demonstrated in studies by Manca and Willan, 2006 and by Manca, et al., 2007. Manca, et al. developed a novel analytical framework that allowed for the appropriate quantification of country-specific cost-effectiveness estimates using individual patient data from multinational RCTs. Their methodology (using Bivariate Hierarchical Linear Model, BHLM) included subject and country-specific covariates to model costs-effectiveness results. They illustrated this method using data from the Atlas trial, a multinational trial that enrolled 3,164 patients in 19 countries comparing low and high doses of the ACE inhibitor, lisinopril, in patients with chronic heart failure. Their country-specific covariates were life expectancy, private and public expenditure on health care as a percentage of GDP, alcohol and tobacco consumption. The selection of these covariates relied on researchers' judgement as non country-specific variables were collected in the trial (usually not done in most of the studies). As patient-level covariates they used age, sex and left ventricular ejection function. By including both country and patient covariates in the model, the authors could explain the between-country variability in the estimated difference in survival. Their results showed that, on average, a patient who was one year older than the overall trial mean would live 4.6 days less than the average patient level in the study. At the country level, those countries with higher (than the overall mean) public expenditure on health care as a percentage of GDP could expect, on average, a survival gain of 52.6 days in patients treated with high-dose lisinopril. This positive and statistically significant correlation between the two variables, expenditure in public healthcare and absolute treatment effects (expressed as the difference in survival), is illustrated in Figure 4.1 from Manca and Willan (2006).

Figure 4.1. Relationship between country-specific difference in survival gain and the difference between the country-specific public healthcare expenditure



Source: Manca and Willan (2006). Relationship between country-specific difference in survival gain between the treatment and control arms and the difference between the country-specific public healthcare expenditure (as a percentage of national GDP) and the mean from all 17 countries in the ATLAS. Note: The size of the markers is inversely proportional to the standard error of the mean survival gain estimate in each country. The dotted lines represent the mean (95% confidence limits) estimated relationship between the mean survival gain and the mean percentage of national GDP in the ATLAS trial in each country.

The authors argued that statistically significant correlation (which does not imply causality) justified the inclusion of country-specific covariates to estimate country-specific treatment effects to assess the generalisability of the results across countries. However, no further discussion on the thinking behind this result was done. One possible explanation could be that the countries with higher expenditure in public health care than the overall mean also have higher baseline risks and therefore benefit the most from the innovative intervention. Unfortunately, the authors pointed out there is very little quantitative research explaining variations in treatment effects across countries; they suggest formulating hypothesis and testing them in practice. Finally, the results of these tests will be also susceptible to the IPD data used, the countries included in the analysis and the variables at country level used to explain the relationship between the absolute treatment effect and country covariates¹².

In a further analysis, the authors tested the differences in the variability of country-specific survival estimates using two different models illustrated in Figure 4.2: The results in Figure 4.2a (Nonhierarchical Bivariate Model) do not use patient and country-specific covariates while results in Figure 4.2b (Bivariate Hierarchical Model, BGLM) control for both types of covariates. Empty square markers indicate country-specific mean differential in survival (high vs low dose) and the overall

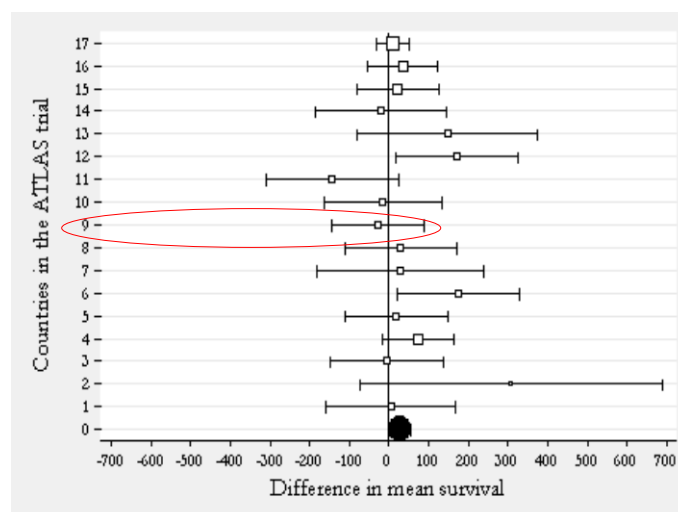
¹² Personal communication with the author.

mean is represented by a black circle marker. The horizontal bars represent 95% credibility intervals. By introducing patient and country-specific covariates (Figure 4.2b), the confidence intervals narrowed significantly reducing the between-country variability in the differential survival (confidence intervals in Figure 4.2b smaller than in Figure 4.2a). Therefore, the proposed hierarchical model produced country-specific estimates closer to the population mean. Also, for some countries, survival estimates changed significantly; results for countries seven and 14 in Figure 4.2b were different from those obtained in Figure 4.2a (using the BLM without covariates), suggesting that low-dose lisinopril dominated high-dose lisinopril. Similarly, results suggest that in countries nine and 10, high dose no longer dominated low dose.

Using this case study, the authors showed that standard cost-effectiveness analysis using individual patient data from multinational trials displayed a large degree of variability across the 17 countries analysed, potentially producing misleading results. Their proposed methods enable more appropriate quantification of country-specific cost-effectiveness estimates by weighting the results based on the level of information available within each country. The consequences are to reduce the degree of variability across countries.

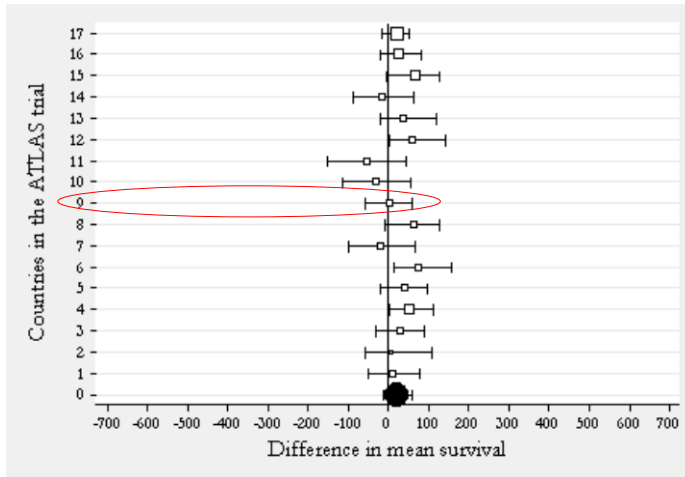
Figure 4.2. Modelled survival gains in the ATLAS trial

(a) Results from the Nonhierarchical Bivariate Model or ‘splitting model’ (not controlling for country or patient covariates)



Source: Manca and Willan (2006). Note: The black circle indicates the trial-wide estimate and the square markers the country-level mean estimates of survival gains. The size of the markers is proportional to the sample of patients recruited in each country. The horizontal bars represent the 95% confidence intervals.

(b) Results from the Bivariate Hierarchical Model (BHLM) controlling for patient and country-specific covariates



The method proposed by Manca, et al. (2007) is only applicable when individual patient data are available and the country for which the analysts are interested in estimating its specific cost-effectiveness ratio participate in the trial. In order to produce unbiased results, as other statistical techniques, centres need to be chosen at random to represent the whole population. This condition is not always fulfilled by trial design.

In cases where this method could not be applied, Manca and Willan (2006) proposed an algorithm to assist with the choice of the appropriate analytical strategy when facing the task of adapting the study results from one country to another. The algorithm considered different scenarios characterised by: (a) whether the country of interest participated in the trial, and (b) whether individual patient-level data (IPD) from the trial are available. The authors emphasised that the methodology used for the generalisability of clinical and economic data in multinational trials is developing rapidly and new options will appear constantly. In some situations, decision modelling studies are the only available option as, for example, when IPD data are unavailable and the country of interest is not in the trial. The authors illustrated this case using a modelling study done by Palmer, et al., (2005) to model the cost-effectiveness of using glycoprotein IIb/IIIa antagonists (GPAs) in the management of non-ST-elevation ACS in the UK. The model was fed by information from international trials but adjustments were required to fit it in the UK setting. Accounted for were differences in clinical practice, in baseline risk because of epidemiology, and in the management of patients. More importantly, Palmer, et al. tested the relationship between the baseline risk and the relative treatment effect or relative risk reduction (RRR), which is usually assumed to be independent and therefore transferable across countries. The results from running a meta-regression using countries' baseline risk and relative risk as estimated in the literature showed a negative but not statistically significant relationship. The model also was adjusted by incorporating UK-specific resource use and costs.

Finally, Vemer, et al. (2010) conducted a modelling study that explored factors potentially driving differences in cost-effectiveness results for three Smoking Cessation Therapies (SCT) -- nicotine replacement therapy (NRT), bupropion, and varenicline plus unaided cessation arm -- across six European countries: The Netherlands (reference case), Belgium, Germany, Sweden, UK and France. The authors attempted to answer the question raised by The Transferability of Economic Data Task Force from ISPOR '[w]hich elements of economic data vary most from setting to setting?' in the SCTs. The nine factors studied were: demography, smoking prevalence, all-cause mortality, smoking-related disease epidemiology, costs of smoking-related diseases, resources used for SCTs, unit costs of SCTs, utility weights and discount rates. The authors explored the impact of between-country differences of these nine parameters on the incremental net monetary benefit (INMB), assuming a threshold of €20,000 per QALY. Unfortunately, they did not report disaggregated results, in particular the impact on QALYs, which would have provided us with evidence exclusively on effectiveness. In order to produce more accurate country-specific clinical data, the model estimated the population baseline risk in each of the six countries although the relative risk reduction was assumed to be generalisable across locations.

Their results showed that the discount rate was the factor contributing the most to the between-country differences in cost-effectiveness, followed by the incidence and mortality of smoking-related diseases and the utility values used to calculate QALYs. These results, however, were sensitive to the assumed value of the cost-effectiveness threshold per QALY, as it affected the estimate of incremental net monetary benefit which valued QALYs at the threshold rate. Their conclusions highlighted the importance of measuring to what extent, and if so how much, the variation in the analysed factors causes variation in cost-effectiveness, as 'the factors that cause the most variation in cost-effectiveness do not necessarily have to be the same as the factors that vary most themselves'.

4.4 Findings and implications

Literature on heterogeneity of relative effectiveness and efficacy across European countries is very scarce. More evidence in the field is needed before any clear statement can be made about the existence of variations in relative effectiveness or efficacy in different countries and in different disease areas. The well accepted assumption of the generalisability of relative treatment effects remains an empirical matter.

All except one of the included studies were cost-effectiveness analyses as in this area; and extensive literature exists dedicated to the heterogeneity and generalisability of results across countries. From those, data on effectiveness or efficacy were extracted. However, the scope of this literature review was not to conduct a systematic review of cost-effectiveness studies as (a) this has been done already (Sculpher, et al., 2004) and (b) our focus is on efficacy and effectiveness.

Heterogeneity in relative *efficacy* across EU countries has not been on the research agenda and therefore not explored in multicountry RCTs. The studies assume clinical effects are homogeneous across patients with similar clinical characteristics regardless of the country in which they receive treatment. Retrospective studies analysing efficacy using individual patient data from multicountry

RCTs have not found statistically significant differences in efficacy either. However, some authors argued that the lack of statistical power of some of the analyses could not detect heterogeneity of efficacy even if present. Some statistical limitations faced in detecting variations in relative efficacy should be overcome in future if new methods keep developing.

More literature was found regarding potential causes of heterogeneity in cost-effectiveness results in Europe. This review focused on those that could be closely related to effectiveness rather than costs. In their review, Sculpher, et al. (2004) identified variation in clinical practice as one of the most cited factors driving differences in cost-effectiveness. However, its impact was mainly used resource consumption to estimate differences in costs and not efficacy or effectiveness.

Vemer, et al. (2010) found the country-specific baseline risk expressed in terms of incidence and mortality of smoking related diseases was an important explanatory factor of the variation of cost-effectiveness results across the analysed countries, contributing to between-country differences in cost-effectiveness results. Manca and Willan (2006) and Manca, et al. (2007) found larger absolute treatment effects (measured as differences in the survival gains based on results from the ATLAS trial) in countries with higher expenditure (as percentage of GDP) in public health care. This finding provides some evidence of country-specific institutional features that could drive variations in absolute treatment effects across countries. However, any relationship to relative treatment effects remains unknown. The general working assumption remains: that relative efficacy (in terms of relative risk reduction) is likely to be constant across countries, but that absolute effects will vary because of differences in baseline risk. For an early discussion of this issue see Jönsson and Weinstein (1997).

Our literature search did not find any observational study analysing heterogeneity in relative *effectiveness* across European countries. However, we are aware that some additional analysis is (has) been done at disease-specific level that can help us understand whether and why differences in relative effectiveness might occur across Europe. Some studies have explored this for cardiovascular disease. One such example is Svilaas, et al., (forthcoming), where GPs in Belgium and Norway were surveyed to ascertain which factors perceived by them affect their ability to evaluate and manage cardiovascular disease risk, including assessment, follow up and communication methods, reimbursement conditions, attitudes towards guidelines, and use of data and tools. One key message from this work is that clinical guidelines and reimbursement conditions may interact to influence prescribing behaviour and treatment options. Goldenberg and Glueck (2009) review the literature on how efficacy data from cholesterol lowering with 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitors in reducing vascular events significantly in primary and secondary prevention of CHD translate into real-life practice. These authors conclude that, in the clinical setting, the rate of statin discontinuation is very high and overall adherence to the medical regimen is low. It is not clear how these factors differ across European countries, but this merits further research.

Evidence is needed as heterogeneity of relative effectiveness across locations is more likely to occur (between and within countries) than relative efficacy. Studies to collect this evidence face several methodological issues and harmonization of criteria is essential to make any comparison. Differences in health outcome measures (clinical endpoints, markers, etc.), patient clinical and socio-

demographic characteristics, and comparators, among others, pose difficulties for any attempt at evaluation. Future collaboration among the Member States, harmonising methodology, generating data and sharing and patients' registries will be crucial to produce this evidence.

4.5 References

Barbieri, M., Drummond, M., Willke, R., Chancellor, J., Jolain, B. and Towse, A. (2005) Variability of cost-effectiveness estimates for pharmaceuticals in Western Europe: Lessons for inferring generalizability. *Value in Health*. 8(1), 10-23.

Bjerrum, L., Munck, A., Gahrn-Hansen, B., Plejdrup Hansen, M., Llor, C., Cots, J.M., et al. (2010) Health alliance for prudent prescribing, yield and use of antimicrobial drugs in the treatment of respiratory tract infections (HAPPY AUDIT). *BMC Family Practice*. 11, 29.

Cook, J.R., Drummond, M., Glick, H. and Heyse, J.F. (2003) Assessing the appropriateness of combining economic data from multinational clinical trials. *Statistics in Medicine*. 22(12), 1955–1976.

Caro, J.J., Huybrechts, K.F., De Backer, G., De Bacquer, D. and Closon, M-C. (for the WOSCOPS Economic Analysis Group). (2000) Are the WOSCOPS clinical and economic findings generalizable to other populations? A case study for Belgium. *Acta Cardiologica*. 55, 239-246.

Drummond, M.F. (1994) Comparing cost-effectiveness across countries. The model of acid-related disease. *Pharmacoeconomics*, 5(Suppl 3), 60-67.

Drummond, M., Barbieri, M., Cook, J., Glick, H.A., Lis, J., Malik, F., Reed, S.D., Rutten, F., Sculpher, M. and Severens, J. (2009) Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. *Value in Health*. 12(4), 409-418.

EUnetHTA. (2007c) Work Package 5. EUnetHTA HTA Adaptation Terms v. 2 and Glossary of HTA Adaptation Terms. Available at http://www.eunetha.net/upload/Barcelona_2007/AdaptationToolkit%20%20Glossary%20combine dJu e2007.pdf [Accessed 10 August 2010].

Goldenberg, N. and Glueck, C. (2009) Efficacy, effectiveness and real life goal attainment of statins in managing cardiovascular risk. *Vascular Health Risk Management*. 5, 369–376.

Hakkaart-van Roijen, L., Verboom, P., Redekop, W.K., Touw, K.R. and Rutten, F.F. (2001) The cost effectiveness of tapered versus abrupt discontinuation of oral cyclosporin microemulsion for the treatment of psoriasis. *Pharmacoeconomics*. 19(5), 599–608.

High Level Pharmaceutical Forum. *Core principles on relative effectiveness*. Available at http://ec.europa.eu/pharmaforum/docs/rea_principles_en.pdf [Accessed 29 July 2010].

Jönsson, B. (2009) *European collaboration on collection and sharing drug effectiveness data*. Conference Presentation at the Swedish EU Meeting, Assessing Drug Effectiveness– Common Opportunities and Challenges for Europe. Stockholm. 29 July.

Jönsson, B. and Weinstein, M.C. (1997) Economic evaluation alongside multinational clinical trials: Study Considerations for GUSTO iib. *International Journal of Technology Assessment in Health Care*. 13, 49-58.

Manca, A. and Willan, A.R. (2006) 'Lost in translation': Accounting for between-country differences in the analysis of multinational cost effectiveness data. *Pharmacoeconomics*, 24(11), 1101–1119.

Manca, A., Lambert, P., Sculpher, M. and Rice, N. (2007) Cost-effectiveness analysis using data from multinational trials: The use of bivariate hierarchical modeling. *Medical Decision Making*. 27(4), 471-490.

Manning, R. and Claxton, K. (1997) *Experimental and econometric solutions to selection bias*. Paper presented to the Health Economics Study Group. York: University of York.

O'Connell, D., Glasziou, P., Hill, S., Sarunac, J., Lowe, J. and Henry, D. (2001) Results of clinical trials and systematic trials: To whom do they apply? In Stevens, A., Abrams, K., Brazier, R., Fitzpatrick, R. and Lilford, R. eds. *The advanced handbook of methods in evidence based healthcare*. London: Sage.

Palmer, S., Sculpher, M., Philips, Z., Robinson, M., Ginnelly, L., Bakhai, A., Abrams, K., Cooper, N., Packham, C., Alfakih, K., Hall, A. and Gray, D. (2005) Management of non-ST- elevation acute coronary syndromes: How cost-effective are IIb/IIIa antagonists in the UK National Health Service? *International Journal of Cardiology*. 100(2), 229-240.

Pinto, E.M., Willan, A.R. and O'Brien, B.J. (2005) Cost-effectiveness analysis for multinational clinical trials. *Statistics in Medicine*. 24, 1965-1982.

Sculpher, M.J., Pang, F.S., Manca, A., Drummond, M.F., Golder, S., Urdahl, H., Davies, L.M. and Eastwood, A. (2004) Generalisability in economic evaluation studies in healthcare: A review and case studies. *Health Technology Assessment*. 8(49)

Svilaas, A., McGrath, C., van Campenhout, H., Grönsleth, M., Fournie, X., Kramarz, P. and De Backer, G.A. (forthcoming) *Descriptive study of cardiovascular risk management and statin therapy in Belgian and Norwegian general practice*.

Vemer, P., Mauren, P.H.M. and Rutten-Van Molken, P. (2010) Crossing borders: Factors affecting differences in cost-effectiveness of smoking cessation interventions between European countries. *Value in Health*. 13(2), 230-241.

Willke, R.D., Glick, H.A., Polsky, D., and Schulman, K. (1998) Estimating country-specific cost-effectiveness from multinational clinical trials. *Health Economics*. 7(6), 481-493.

Zink, A., Askling, J., Dixon, W. G., Klareskog, L., Silman, A.J. and Symmons, D.P.M. (2009) European biological registers: Methodology, selected results and perspectives. *Annals of Rheumatic Diseases*. 68, 1240-1246.

5 Putting the Jigsaw Together with an Industry Vision of the Future

5.1 Introduction

The policy issue is no longer whether or not the pharmaceutical industry should support moves for a pan-European assessment of Relative Efficacy and/or Relative Effectiveness; it is whether or not a pan-European process can be designed which meets the needs of the industry and how the industry should engage to bring this about.

5.2 Summarising our findings

The position as set out in Chapters 2 to 4 of our report can be summarised as follows.

Analysis of what Member States are currently doing suggests that a pan-European approach to relative efficacy might be possible as all of the bodies we looked at appear to use this information. However, at least three issues that can affect variation in relative efficacy need highlighting: (1) selection/exclusion of studies (RCTs in particular) to include in the assessments; (2) selection of end points for consideration; and (3) methodology used for reviewing RCTs. For instance, one issue will be the use of indirect comparisons where head-to-head RCTs do not exist. Of course, if there are differences in existing clinical practice between Member States then the choice of comparator may be different. Evidence on this is limited without exploring in detail assessments of specific drugs across countries. We have included such an assessment in the future research agenda.

On the translation of relative efficacy into relative effectiveness, important differences appear to exist across countries. One group of countries appears to use relative efficacy evidence to determine degree of innovativeness which is then the basis for reimbursement decisions and pricing negotiations. Others seek to move from efficacy to an estimate of effectiveness, but use different approaches. For some, it is identifying more patient centred outcomes, or longer term effects; for others, it is understanding whether the RCT results are likely to be replicated in routine use.

Looking at the current pan-European processes underway, it is clear that EUnetHTA is making the most 'noise', undertaking a second programme of work with financing from the European Commission. However, the results of the first programme of work were much more focused on principles than on practicalities. The 'core model' worked example of DES versus BMS was one of the few practical outputs that enables EFPIA to begin to understand whether the collaborating groups do understand the complexity of the process in which they are engaged. The results need to be reviewed carefully but, from our initial reading, they give little indication as to how the new WP5 on the Relative Effectiveness of Pharmaceuticals might be handled.

Our understanding of the literature on heterogeneity of relative effectiveness and efficacy across European countries is that:

- Heterogeneity in relative efficacy across EU countries has not been on the research agenda and therefore not explored in multi-country RCTs. The studies assume clinical effects are

homogeneous across patients with similar clinical characteristics regardless of the country in which they receive treatment. Retrospective studies analysing efficacy using individual patient data from multi-country RCTs have not found statistically significant differences in efficacy either. However, some authors argue that the lack of statistical power of some of the analyses could not detect heterogeneity of efficacy even if present. Some of these statistical limitations to detecting variations in relative efficacy will be hopefully overcome in future if new methods keep developing.

- More literature was found regarding potential causes of heterogeneity related to effectiveness. Variation in clinical practice was one of the most cited factors driving differences in cost-effectiveness. However, its impact was mainly assessed based on resource consumption to estimate differences in costs and not in efficacy or effectiveness, leaving open the question as to how important these differences were for relative effectiveness as opposed to cost-effectiveness, which is quite different. More evidence in the field is needed before any statement can be made about the existence (or not) of variations in relative effectiveness in different countries.
- The search did not find any observational study analysing heterogeneity in relative effectiveness across European countries. Evidence is needed as heterogeneity of relative effectiveness across locations is more likely to occur (between and within countries).

5.3 What could be the benefits of pan-European assessment?

A pan-European assessment of relative efficacy would avoid duplication of effort, and depending on how it was done and by whom, could raise the quality of assessments in some Member States. However, it is not clear that any time would be saved. The EUnetHTA work to date has given little indication of the practical issues that might arise. It may make sense for the industry to engage with MEDEV and other payers who are collaborating to get a better sense of how specific issues are being dealt with, as well as getting involved in the EUnetHTA work.

The benefits to the industry of a pan-European assessment would depend on the methods used to assess the RCT evidence, particularly on the use of indirect comparators. In a worst case, the assessors may conclude that there is no evidence that a new product is better than a comparator because they are not prepared to accept an indirect comparison as valid and a head-to-head RCT does not exist. We return to this point of achieving realistic expectations below.

5.4 Relative effectiveness and the post-launch environment

Differences between countries as to whether and how they assess relative effectiveness clearly exist. However, this is separate from the question as to whether the relative effectiveness of drugs is likely to vary across countries for reasons that are independent of the choice of comparator. Possible reasons why this might happen include differences in population composition and differences in the baseline risk (which, of course, is linked to population heterogeneity.)

The industry's model for the EU (and the US) is likely (we speculate) to be one of streamlined development, i.e. an attempt to shift any additional data requirements from a pre-launch to a post-launch environment where companies get some sort of revenue stream. Companies may of course choose to do more pre-launch to avoid post-launch burdens, improve their competitive position at launch, and eliminate any conditionality to marketing authorisation or to reimbursement /coverage, but they are likely to want this to be a commercial decision rather than one imposed by marketing authorisation bodies. Although demonstrating value at launch will be important, companies will want the opportunity to offer forms of 'coverage with evidence development'. Willingness to enter into such arrangements is likely to vary by Member State. However, when companies do undertake post launch studies of effectiveness, for whatever reason, they are likely to want the results to be generalisable. If there is an a priori assumption that effectiveness is not transferable, then companies may be pushed into unnecessary additional studies.

To this end, the efforts initiated by the Swedish Presidency that we discuss in Section 3 of the report, intended to encourage pan-European use of registries and observational data collection, are very important to the industry. One important question (yet unanswered) is who finances the registries and the data collection and who owns the data generated by these initiatives.

5.5 The dangers of overly high standards

Relative efficacy and relative effectiveness may well vary across Member States. The choice of comparator issue is very important, particularly in areas such as oncology where many treatment combinations are possible and clinical practice is very likely to vary from one Member State to another. Even where the comparator is the same, relative effectiveness may differ (whilst relative efficacy does not) because of other aspects of clinical practice.

It will not be straightforward to get a common approach from HTA/P&R bodies to relative efficacy and relative effectiveness, particularly the larger ones (which are in the more important markets). It is certainly realistic to try to get more commonality in methods and data requirements for assessing relative efficacy and relative effectiveness, beginning with relative efficacy. It is important, however, that the effect of such an initiative does not lead to the 'highest common denominator' in terms of restrictive methods and excessive data requirements. Engagement with key HTA/P&R bodies independently of specific EUnetHTA projects may make sense. Two methodological areas where this dialogue will be particularly important include (a) the use of indirect comparisons when head to head comparisons are not available and (b) observational study design to collect data for relative effectiveness purposes.

The EMA clearly has expertise to contribute to such a dialogue around relative efficacy, and indeed is likely to take a more realistic view on issues of concern to industry such as the feasibility of 'head-to-head' trials in the development process. Indeed if industry were successful in moving to earlier launch and earlier access with post-launch data requirements both for EMA and HTA/P&R bodies (forms of conditional authorisation and/or conditional coverage), then some sort of tripartite scientific dialogue around *post-launch* data collection may even begin to make sense with EMA playing a role in this. This possibility would be reinforced by the growing post-launch benefit-risk

role of the EMA. This involves understanding benefits post-launch, as well as risks. Both require observational data. The EFPIA decision is not to seek any increase in the mandate of the EMA but to welcome its contribution under the existing mandate, including greater clarity in the EPAR to enable HTA/P&R bodies to have a better understanding of both the evidence the EMA has reviewed and the conclusions it has drawn. It will be important that the industry ensure that EMA is involved actively in dialogue around pan-European assessment of relative efficacy not because it is a candidate to do the analysis, but because its expertise is crucial to helping the industry get realistic approaches to a pan-European assessment of relative efficacy.

An issue that warrants further consideration is understanding options for *process* in any pan-European assessment. For example, the model could be that a single Joint Action country takes the lead on the relative efficacy assessment that is then adopted by other countries within the Joint Action as a starting point. Whether this is positive or negative depends on the standard and methodology being adopted. This should be agreed up front, without letting the 'lead country' apply its own approach without a consensus.

Appendix 1. Current Use of Relative Effectiveness in Selected EU Member States

England and Wales

Pricing and Reimbursement System: General Overview

In England and Wales, the Department of Health relies on an arms-length Special Health Authority, the National Institute for Health and Clinical Excellence (NICE), to make recommendations on adoption of pharmaceutical (and other) health technologies. NICE's recommendations are based on evidence submitted by key stakeholders e.g. manufacturers and independent assessment centres.

In making adoption decisions, NICE employs a cost effectiveness threshold range of £20,000 to £30,000 per QALY. This 'threshold' was explicitly formalised for the first time in 2004 (NICE, 2004). There is however some flexibility attached to this threshold range and health technologies may be recommended for reimbursement if the cost per QALY falls outside this range once a number of additional criteria (including 'innovativeness', the impact on wider social costs, the nature of the disease and the severity of the ill health of the target patient population) are taken into account.

The Institute has two appraisal processes for technology appraisals: the multiple technology appraisal (MTA) process and the single technology appraisal (STA) process. Although there are differences between the two processes, the principles relating to decision-making and the methods of assessment are the same.

The appraisal of a health technology is divided into three distinct phases.

- Scoping
- Assessment
- Appraisal

Scoping: The questions to be addressed by the appraisal are fundamental to the assessment process and require an understanding of the context within which a technology is to be investigated, including currently available care and any alternative technologies for the specific indication.

Assessment: The assessment process is a systematic evaluation of the relevant evidence available on a technology. The aim is to produce an estimate, taking account of uncertainty, of a technology's clinical- and cost-effectiveness for a specific indication. Assessment normally has two mutually dependent components: a systematic review of the evidence and an economic evaluation. The assessment process always includes a review of the evidence by an independent assessment group.

Appraisal: The appraisal process is a consideration of the reports and analyses produced in the assessment phase within the context of additional information supplied by consultees, commentators, clinical specialists, patient experts and the general public. The Appraisal Committee considers the evidence available and then formulates an appraisal decision, applying judgements on the importance of a range of factors that may differ from appraisal to appraisal.

The Secretary of State for Health has directed that the NHS provides funding and resources for technologies that have been recommended through the NICE technology appraisals programme normally within three months from the date that the guidance is published.

Guidelines

Analytical method

The 'Reference case' specifies the methods considered by the Institute to be the most appropriate for the Appraisal Committee's purpose and consistent with an NHS objective of maximising health gain from limited resources. Submissions to the Institute should include an analysis of results generated using these reference-case methods. This does not preclude additional analyses being presented when one or more aspects of methods differ from the reference case. However, these must be justified and clearly distinguished from the reference case.

The Table below, which has been extracted from NICE's guidelines, summarises the reference case.

Element of health technology assessment	Reference case	Section providing details
Defining the decision problem	The scope developed by the Institute	5.2.5 & 5.2.6
Comparator	Therapies routinely used in the NHS, including technologies regarded as current best practice	5.2.5 & 5.2.6
Perspective on costs	NHS and PSS	5.2.7 to 5.2.10
Perspective on outcomes	All health effects on individuals	5.2.7 to 5.2.10
Type of economic evaluation	Cost-effectiveness analysis	5.2.11 & 5.2.12
Synthesis of evidence on outcomes	Based on a systematic review	5.3
Measure of health effects	QALYs	5.4
Source of data for measurement of HRQL	Reported directly by patients and/or carers	5.4
Source of preference data for valuation of changes in HRQL	Representative sample of the public	5.4
Discount rate	An annual rate of 3.5% on both costs and health effects	5.6
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	5.12

HRQL, health-related quality of life; NHS, National Health Service; PSS, personal social services; QALYs, quality-adjusted life years.

For the reference case, cost-effectiveness (specifically cost-utility) analysis is the preferred form of economic evaluation. This seeks to establish whether differences in costs between options can be

justified in terms of changes in health effects. Health effects should be expressed in terms of quality-adjusted life year (QALYs).

Health outcomes/ QoL measure

As far as possible, principal measures of health outcome are identified in the scope. For the valid analysis of clinical effectiveness, the principal outcome(s) should be clinically relevant; that is, they measure health benefits and adverse effects that are important to patients and/or their carers. The clinical outcome measures would usually be expected to have an impact on survival or health-related quality of life (HRQL) and be able to be translated into quality-adjusted life years (QALYs) for the evaluation of cost effectiveness.

The requirements for evidence of effectiveness include the quantification of the effect of the technologies on disease progression and patients' HRQL, and the valuation of those effects in a manner that reflects the preferences of the general population.

The Institute has a strong preference for expressing health gains in terms of QALYs. In most circumstances, when the health gain is expressed in terms of life-years gained, the range of most plausible 'life-years gained' ICERs that are acceptable will be substantially lower than those described above (£20,000-£30,000). In these circumstances, the Committee will impute a plausible QALY value from the estimated life-years gained. The exact adjustment that the Committee makes will take account of the differences between QALYs and life-years gained. It will be guided by reference to the population norms for HRQL for the affected population, but will generally be lower than this for a sick population.

For cost-effectiveness analysis, the value of health effects should be expressed in terms of QALYs for the appropriate time horizon. For the reference case, the measurement of changes in HRQL should be reported directly from patients and the value of changes in patients' HRQL (that is, utilities) should be based on public preferences using a choice-based method. The EQ-5D is the preferred measure of HRQL in adults. The methods to elicit EQ-5D utility values should be fully described. When EQ-5D data are not available or are inappropriate for the condition or effects of treatment, the valuation methods should be fully described and comparable to those used for the EQ-5D. Data collected using condition-specific, preference-based measures may be presented in separate analyses. The use of utility estimates from published literature must be supported by evidence that demonstrates that they have been identified and selected systematically.

In the reference case, an additional QALY will receive the same weight regardless of any other characteristics of the people receiving the health benefit.

Perspective

For the reference case, the perspective on outcomes should be all direct health effects, whether for patients or, when relevant, other people (principally carers). The perspective adopted on costs should be that of the NHS and personal social services (PSS). Technologies for which a substantial proportion of the costs (or cost savings) are expected to be incurred outside of the NHS and PSS, or which are associated with significant non-resource effects other than health, should be identified during the scoping stage of an appraisal. In these exceptional circumstances, information on costs to

other government bodies, when these are not reflected in HRQL measures, may be reported separately from the reference-case analysis. The intention to include such data will normally be agreed with the Department of Health before finalisation of the remit.

Evidence considered

Consideration of a comprehensive evidence base is fundamental to the appraisal process. Evidence of various types and from multiple sources may be relevant to the appraisal. The guiding principle around the evidence is that all relevant evidence of the best available quality needs to be assembled systematically and synthesised in a transparent and reproducible manner.

The guidelines broadly categorise the primary research methods and designs used to measure the treatment effect as experimental or observational studies. The guidelines consider experimental studies with high internal and external validity as the most reliable evidence about the relative treatment effects of a technology. For an assessment of internal validity, the different types of study design can be ranked according to design features that affect their validity for estimating relative treatment effect, ranging from RCTs to uncontrolled observational studies.

NICE has a strong preference for evidence from 'head-to-head' RCTs that directly compare the technology with the appropriate comparator in the relevant patient groups. When such evidence is available and includes relevant outcome evidence, this is preferred over other study designs. The rationale for the identification and selection of the RCTs should be explained, including the rationale for the selection of treatment comparisons that have been included. A clear description of the methods of synthesis is required.

But the guidelines do recognise the limitations around RCTs – including limited selected populations and comparator treatments and short time spans that do not reflect routine or best NHS practice. Therefore, it is argued that good-quality non-randomised studies, both experimental and observational, may be needed to supplement RCT data. This might be the case when there is a need to estimate relative treatment effect over longer time horizons or to measure particular outcomes that have not been included in the RCTs. But the guidelines state that inferences about relative treatment effects drawn from non-RCT evidence will necessarily be more circumspect than those from RCTs with properly controlled. Moreover, any potential bias arising from the design of the studies used in the assessment should be explored and documented.

When head-to-head RCTs exist, evidence from mixed treatment comparison analyses may be presented if it is considered to add information that is not available from the head-to-head comparison. This mixed treatment comparison must be fully described and presented as additional to the reference-case analysis (a 'mixed treatment comparison' includes trials that compare the interventions head-to-head and indirectly). When multiple technologies are being appraised that have not been compared within a single RCT, data from a series of pairwise head-to-head RCTs should be presented. Consideration should also be given to presenting a combined analysis using a mixed treatment comparison framework if it is considered to add information that is not available from the head-to-head comparison.

If data from head-to-head RCTs are not available, indirect treatment comparison methods should be used (an 'indirect comparison' is defined therein as a synthesis of data from a network of trials). The principles of good practice for standard meta-analyses should also be followed in mixed and indirect treatment comparisons. When this is the case, however, the analysis may be restricted to a qualitative overview that critically appraises individual studies and presents their results. In these circumstances, the Appraisal Committee will be particularly cautious when reviewing the results of analysis.

The guidelines also state that the synthesis of outcome data through meta-analysis is appropriate provided there are sufficient relevant and valid data that use measures of outcome that are comparable.

In considering cost effectiveness, the guidelines state that it is likely that evidence from other study designs in addition to that from RCTs will be necessary. Evidence on cost effectiveness may be obtained from new analyses; however, a systematic review of published, relevant evidence on the cost effectiveness of the technology should also be conducted.

The Institute will normally be supplied with evidence from:

- An independent assessment group (the 'Assessment Group' for MTAs and the 'Evidence Review Group' for STAs)
- Manufacturers and sponsors of technologies
- National patient/carer groups
- Healthcare professionals
- Clinical specialists and patient experts

In the MTA process, the Assessment Group prepares the assessment report, which is an independent synthesis of the evidence from published information and the submissions from manufacturers and sponsors about the clinical- and cost-effectiveness of the technology/technologies. The report provides a systematic review of the literature and a review of manufacturer and sponsor economic models submitted to the Institute. It usually includes a new assessment of cost effectiveness based on an economic model. The Assessment Group also consults clinical and methodological experts, and patient groups, when gathering evidence for the assessment report.

In the STA process, the Evidence Review Group prepares the Evidence Review Group report, which is a critical appraisal of the submission provided by the manufacturer or sponsor of the technology. If the Evidence Review Group is concerned about any assumptions made in the submitted analyses, it may recommend that the Institute requests additional analysis from the manufacturer or sponsor, and/or may undertake additional analysis themselves.

The guidelines also refer to the evidence that manufacturers should include in their submission, which includes:

- An assessment of clinical effectiveness, containing a critical appraisal, interpretation and synthesis of clinical-effectiveness evidence

- A tabulation of the values and sources of the key parameters to be used in the assessment of cost effectiveness
- An assessment of cost effectiveness containing a reference-case analysis of cost effectiveness based on the synthesis of clinical-effectiveness evidence. A justification for any cost-effectiveness analysis not fulfilling the reference-case requirement is essential

There is also a discussion about the use of evidence that is not in the public domain. NICE recommends the use of this evidence to be kept to a minimum. However, under exceptional circumstances, the Institute will accept unpublished evidence under agreement of confidentiality; for example, if the information is commercially sensitive ('commercial in confidence') or if its use might adversely affect future publication rights ('academic in confidence').

Comparator

The relevant comparators for the technology being appraised are those routinely used in the NHS, and therapies regarded as best practice (including existing NICE guidance) when this differs from routine practice. The guidelines accept that there will often be more than one relevant comparator technology because routine practice may vary across the NHS and because best alternative care may differ from routine NHS practice. For example, this may occur when new technologies are used inconsistently across the NHS. Relevant comparator technologies may also include those that do not have a marketing authorisation for the indication defined in the scope but that are used routinely for the indication in the NHS. Comparator technologies may include branded and non-proprietary (generic) drugs.

Choice of patient group

The guideline acknowledges that for many technologies, the capacity to benefit from treatment will differ for patients with differing characteristics. Thus, this needs to be explored as part of the reference-case analysis by the provision of estimates of clinical- and cost-effectiveness separately for each relevant subgroup of patients. The characteristics of patients in the subgroup should be clearly defined and should preferably be identified on the basis of an a priori expectation of differential clinical or cost effectiveness due to known, biologically plausible mechanisms, social characteristics or other clearly justified factors. When possible, potentially relevant subgroups will be identified at the scoping stage with consideration being given to the rationale for the expectation of a subgroup effect. However, this does not preclude the identification of subgroups later in the process; in particular, during the deliberations of the Appraisal Committee.

Analysis of uncertainty / sensitivity analysis

Overall, the uncertainty surrounding the estimates of clinical- and cost-effectiveness needs to be fully expressed. As stated in the guidelines, the Appraisal Committee needs to be able to fully appreciate the uncertainty and limitations associated with the clinical and cost-effectiveness evidence. Consideration of the uncertainty and limitations of the evidence base is needed to provide a robust evaluation of the expected costs and health effects of a technology, to assess whether existing evidence is sufficient for recommending the routine use of a technology, and to enable consideration of the possible consequences of an uncertain decision for the NHS. This requires the

appropriate use of rigorous methods to assess the implications of uncertainty, including the uncertainty around the appropriate structure of the economic model, the choice of sources and analyses to inform the estimates of costs and health effects, and the precision with which these are known. This quantification of decision uncertainty may then feed into subsequent decisions about the need for future research.

The models used to synthesise available evidence to generate estimates of clinical- and cost-effectiveness for the Institute's needs should follow accepted guidelines. Full documentation and justification of structural assumptions and data inputs should be provided. When there are alternative plausible assumptions and inputs, sensitivity analyses of their effects on model outputs should be undertaken.

Sensitivity analysis should be used to explore uncertainty around the structural assumptions used in the analysis. Analysis of a representative range of plausible scenarios should be presented and each alternative analysis should present separate results.

Costs

Costs to the NHS and the PSS should be reported. These resources should be valued using the prices relevant to the NHS and PSS. Evidence should be presented to demonstrate that resource use and cost data have been identified systematically.

Some technologies may have a substantial impact on the costs (or cost savings) to other government bodies. In these exceptional circumstances, costs to other government bodies may be included if this has been specifically agreed with the Department of Health, usually before referral of the topic. When non-reference-case analyses include these broader costs, explicit methods of valuation are required. In all cases, these costs should be reported separately from NHS/PSS costs. These costs should not be combined into an incremental cost-effectiveness ratio (ICER) where the QALY is the outcome measure of interest.

Time horizon

The time horizon should be sufficient to reflect important cost and benefit differences between the technologies being compared.

Discounting

For the reference case, an annual discount rate of 3.5% should be used for both costs and benefits. When results are potentially sensitive to the discount rate used, consideration should be given to sensitivity analyses that use differential rates for costs and outcomes and/or that vary the rate between 0% and 6%.

Presentation of methods/results

All parameters used to estimate clinical- and cost-effectiveness should be presented clearly in tabular form and include details of data sources. For continuous variables, mean values should be presented and used in the analyses. For all variables, measures of precision should be detailed. For probabilistic analyses, the distributions used to characterise the uncertainty in input parameters

should be documented and justified. As much detail as possible on the data used in the analysis should be provided.

The expected value of each component of cost and expected total costs should be presented; expected QALYs for each option compared in the analysis should also be detailed in terms of their main contributing components. ICERs should be calculated as appropriate.

References

National Institute for Health and Clinical Excellence (NICE). (2008) *Guide to the methods of technology appraisal*. London: National Institute for Health and Clinical Excellence

France

Pricing and Reimbursement System: General Overview

France has one of the longest established and most sophisticated formal pricing and reimbursement systems. It is administered by the *Haute Autorité de Santé* (HAS), using a dual ‘technical – economic’ committee model.

The ‘Improvement in the Medical Service’ (ASMR) I-V rating scale for the contribution of new medicines to the health service is based primarily upon expert clinical rapporteurs and high level clinical committees who create and review evidence in order to make judgements on allocating products across the ASMR range. The Economic/Pricing Committee uses the ASMR and any submissions from companies to make a judgement and if necessary negotiate prices with individual companies.

Prior to assessing a medicine’s ASMR rating, a ‘medical value’ (SMR) rating is established – which determines medicines’ reimbursement status. The SMR rating is based on the following factors:

- Efficacy/tolerance
- Severity of the disease
- Existence of therapeutic alternatives
- Place in the therapeutic strategy (first line, second line, etc.)
- Public health impact

Cost-effectiveness is not a criterion. There are four possible SMR levels: major or important; moderate; low; or insufficient. It is usually ‘efficacy’ and ‘disease severity’ that determine the medicine’s SMR classification. Table 1 shows the relationship between SMR, severity of disease and reimbursement.

Table 1. Relationship between SMR rating, severity of disease and patient co-payment rate

SMR	Disease	
	Severe	Non-severe
Major or important	35 %	65 %
Moderate or low	65 %	65 %
Insufficient	100 %	100 %

The ASMR criterion is more dynamic, since it relates to a medicine’s therapeutic value relative to the existing situation/available treatments. There are five possible grades to this classification:

- Major (Level I) - ‘Major therapeutic progress’;
- Important (II) - ‘Important improvement in terms of therapeutic efficacy and/or tolerance’;
- Modest (III) - ‘Modest improvement in terms of therapeutic efficacy and/or tolerance’;
- Minor (IV) - ‘Minor improvement in terms of therapeutic efficacy and/or tolerance’; and
- No improvement (V) - ‘Absence of improvement with unfavourable appraisal for inclusion on positive list’.

Figures 1 and 2 show the number of medicines being given each of the SMR and ASMR ratings respectively by the HAS in 2008.

Figure 1. SMR ratings

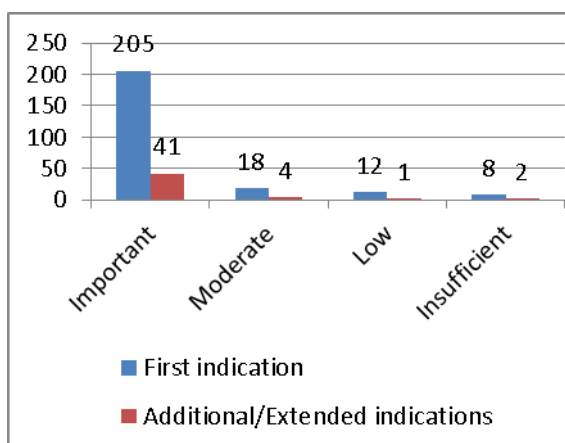
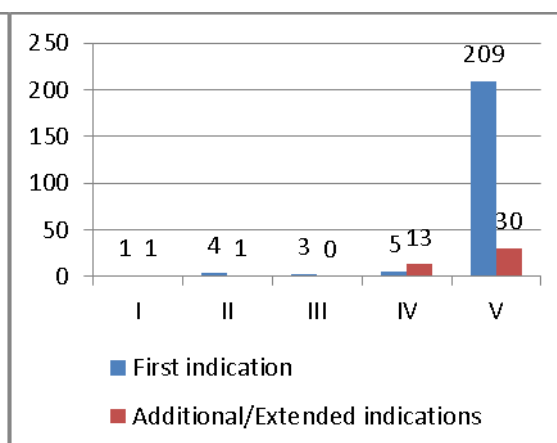


Figure 2. ASMR ratings



Source: HAS 2008 Annual Report (available, in French, at: http://www.has-sante.fr/portail/upload/docs/application/pdf/2009-07/rapport_activite_2008.pdf [Accessed 26 April 2010]). NB Each decision can include more than one SMR/ASMR rating.

As shown in Figure 1, 205 ratings out of a total of 243 SMR ratings by the HAS (84%) for first indications (blue bars) were 'Important' in absolute terms in 2008. However, and as shown in Figure 2, 94% of ASMR ratings for first indications (209 out of 222) were deemed by the HAS to offer 'no improvement' relative to existing therapies (i.e. were given a 'V').

Guidelines

No detailed guidelines were found.

Germany

Pricing and reimbursement system: General overview

In Germany, medicines are automatically reimbursed after marketing authorisation, with the exception of products for minor illnesses and so-called 'lifestyle' drugs.

The degree of reimbursement is defined by a reference price system, which was first introduced in 1989. On-patent drugs have to exhibit in the view of the *Gemeinsamer Bundesausschuss* (G-BA) significant additional therapeutic benefits or fewer serious side effects than existing comparable medicines in order to be excluded from the reference price system. There are three reference price groups: products with the same active ingredient (group 1), products with therapeutically and pharmacologically similar active ingredients (group 2) and compounds with comparable therapeutic effects (group 3). Groups 2 and 3 can include patent-protected drugs as well as off-patent branded and generic products - these mixed groups are termed 'jumbo groups'. Level 2 groups can also be formed with patent protected drugs only (minimum of three drugs). These new groups as currently defined cover, among others: proton pump inhibitors, statins, sartans/AT1 antagonists, triptans, fluoroquinolone antibiotics and alpha-glucosidase inhibitors.

No formal pharmacoeconomic evaluation is required to set reference prices. However, the 2004 health care reform saw the formation of the Institute for Quality and Economic Efficiency in Health Services (*Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen* [IQWiG]), which is responsible for assessing clinical evidence and providing treatment guidelines for key diseases and (since April 2007) for assessing the costs and benefits of pharmaceuticals which are not included in the reference price system. The decisive body, the Joint Federal Committee (*Gemeinsamer Bundesausschuss* [G-BA]), which consists of representatives of Social Health Insurers (*Krankenkassen* [SHI]) and Accredited Office-based Physicians, makes decisions based on the Institute's recommendations on therapeutic guidelines, and on limitations of and/or exclusions from reimbursement of medicines.

In January 2008, IQWiG issued a draft of its 'Methods' as a result of a consultation process with an international Expert Panel (IQWiG, 2008)¹³. This paper mainly focuses on the concept of an efficiency frontier. To compare costs and benefits in a particular therapy area, the paper argues that a diagram should be constructed with 'costs' on the x-axis and 'value' on the y-axis. The idea is then to plot the existing therapies as points on the graph so that comparisons can be made. Thus, where alternative therapies exist, a new medicine may be priced for reimbursement by the SHI no higher than would yield the same cost per unit of outcome as the existing medicine. Units of outcome used are disease specific. In those situations where no alternative therapy exists, the new medicine would be the first point to plot onto the efficiency frontier graph and so this method cannot be used to determine the reimbursed price¹⁴.

As a basic principle, Statutory Health Insurance (SHI) insurants may not be deprived of access to beneficial health technologies on cost grounds alone. In consequence, effective therapies are initially adopted regardless of price. Recognizing that this approach will not be permanently sustainable, the legislation now provides an instrument to define the maximum reimbursable price, fixing the limit up to which health insurance funds can reimburse costs. For this purpose, IQWiG has developed a method for the health economic evaluation of drugs and other interventions. Assessments of the relation of benefits to costs based on this method are then submitted to the G-BA as recommendations for decision making.

Guidelines

Analytical method – economic model

As outlined above, IQWiG's analytical method comprises of creating an 'efficiency frontier'. An efficiency frontier is constructed for each therapeutic area as the basis for health economic evaluation of relevant health technologies. But to undergo a health economic evaluation, health technologies have to possess an additional benefit compared to other health technologies already available or other therapeutic alternatives in use in the health care system.

Without employing a universal threshold – which is currently non-existent in Germany – the efficiency frontier method is based on the determination of the prevailing efficiency in a given therapeutic area in Germany. The efficiency frontier itself is comprised of the most efficient therapeutic alternatives within the particular therapeutic area. Recommended actions for the decision maker can be derived from the last plotted point on the efficiency frontier (technology showing the highest benefit).

¹³ IQWiG (2008),. Methods for Assessment of the Relation of Benefits to Costs in the German Statutory Health Care System, version 1.0

¹⁴ As of 16 April, there were two ongoing health economics evaluations: 'Health economic evaluation of venlafaxine, duloxetine, bupropion, and mirtazapine in depression' and 'Health economic evaluation of clopidogrel in acute coronary syndrome and in peripheral occlusive arterial disease' (<http://www.iqwig.de/index.286.en.html>). We accessed this webpage on the 26 April 2010. As of 29 July 2010, this webpage no longer exists.

The additional therapeutic benefit derived from a previous benefit assessment may be transferred into an approximately cardinally scaled measure. Interventions on the efficiency frontier denote the net cost for any given benefit that is consistent with the efficiency that can be achieved by the package of interventions on the current market. Prices can lead to health technologies being positioned on an already existing segment of the efficiency frontier, showing thereby consistent efficiency with already existing interventions. If a price results in an intervention being positioned below the efficiency frontier, this indicates a lower efficiency. This price is deemed too high and needs to be adjusted, or at least justified. Interventions above the efficiency frontier indicate improved efficiency and thus redefine the frontier.

If health economic evaluations are carried out using different clinical measures, an efficiency frontier needs to be established for each one of these measures. The Methods paper also states that another way of representing the benefit on the benefit axis of the efficiency frontier graph is to aggregate different benefits into one single measure and then to establish one single efficiency frontier.

The specific modelling technique should not be determined in advance. For this reason, IQWiG has no *a priori* preference for a specific modelling technique. The choice of the appropriate modelling technique depends on the research questions embodied in IQWiG's commission from the G-BA, the characteristics of the technologies to be assessed, the disease and the general conditions.

The efficiency frontier plot is designed in such a way that it represents the relevant health technologies in a given therapeutic area. This involves:

- Full, detailed specification of the therapeutic area in question. This may include the specific disease, the conditions of treatment (e.g. inpatient care), target population, ranking of therapy (first, second choice, etc.), and whether it is a mono-therapy or combination therapy
- Scoring existing therapies on the basis of benefits and costs
- Entering therapies on a coordinate system with the benefit on the vertical axis and costs on the horizontal
- Drawing the efficiency frontier

In order to construct the frontier there are three major steps to take:

- Defining the vertical axis, by quantifying the benefit for the chosen interventions and ensuring that an approximately cardinal scale is used which reflects the benefit in the therapeutic area in question (Multiple efficiency frontiers may be derived and can be presented to the decision maker)
- Defining the horizontal axis, by quantifying the total net costs per patient for each of the selected therapies
- Plotting the interventions and draw the efficiency frontier

Medical interventions are compared with other clearly defined active or sham interventions (e.g. placebo), or with no intervention in respect of their (additional) beneficial and harmful effects on defined patient-relevant outcomes. For this purpose, exactly one of the five following evaluating

conclusions is first drawn for each predefined patient-relevant outcome on the basis of the analysed scientific data available:

- Proof of a(n) (additional) benefit or harm exists
- Indications of a(n) (additional) benefit or harm exist
- Proof of the lack of a(n) (additional) benefit or harm exists
- Indications of the lack of a(n) (additional) benefit or harm exist
- No proof and no indication of a(n) (additional) benefit or harm exist

As a rule, if the conclusion is drawn that 'proof' is available, it is required that a meta-analysis of studies shows a corresponding statistically significant effect (with outcome-related minor uncertainty of results). If a meta-analysis is not feasible, at least 2 studies conducted independently of one another should be available that show outcome-related minor uncertainty of results and a statistically significant effect, and whose results are not questioned by further comparable studies with outcome-related sufficient certainty of results ('consistency of results'). The two studies conducted independently of one another need not necessarily be of exactly identical design. Which deviations in design between studies are still acceptable depends on the research question posed. Despite showing statistically significant effects, as a rule a meta-analysis of studies with outcome-related high uncertainty of results or results from individual studies can consequently at most provide indications of the effects of an intervention. If, in exceptional cases, proof of the benefit of an intervention is inferred from only one study, then specific requirements apply to this study and its results.

Health outcomes / QoL measure

The use of QALYs specific to each therapeutic area is not ruled out by IQWiG, but there are ethical and methodological concerns arising from certain survey instruments, such as time trade off and standard gamble, to consider prior to implementation. No universally accepted method has yet been found. Instead, the IQWiG method takes a rather pragmatic approach aimed at comparing the efficiency of treatments in a therapeutic area, without addressing the broader issue of prioritizing across the health care system.

The primary clinical measures used by IQWiG are mortality, morbidity, health-related quality of life and validated surrogates. The following outcomes related to patient benefit are to be given particular consideration: increase in life expectancy, improvement in health status and QoL, as well as reduction in disease duration and adverse effects. These dimensions of benefit are represented by the outcomes listed above (mortality, morbidity and health-related QoL).

Health economic evaluations in Germany are not performed across indications, but only within individual therapeutic areas. Thus, single indication-specific aggregated measures can be employed. It is not necessary to use primarily aggregated measures which can be applied across indications. The use of such measures, e.g. the QALY, can however be reasonable for the comparison of interventions within a therapeutic area, if there is no other validated instrument for aggregating the assessment of benefit and harm in this area. The indication-specific use of QALYs can be particularly

useful with new drugs whose life-extending effect is considerably offset by the reduction of quality of life caused by side effects.

Perspective

As a general rule the perspective must be that of the SHI insurants. Based on this perspective, the disease-related benefits covered by SHI are reflected as well as co-payments by insurants.

Evidence considered

Typically, data will be drawn from several sources, including trials on treatment effects, cohort studies that investigate specific parameters and risk factors related to the natural course of the disease and the associated life expectancy, trans-sectional surveys collecting quality of life data, registry data on resource use and costs as well as automated databases or compiled statistics. Assumptions based on expert opinion should be avoided, as they are rarely accurate enough to be used when other sources of evidence are not available. An important criterion for the studies is the transferability of results to the German context. Due to the differences between systems, transferring cost data from other health systems is seldom possible and, if done, then only under stringent conditions.

The Methods paper states that one of the key findings of EBM is that study types other than RCTs are usually not suited to prove causality. It is possible to investigate patients under everyday conditions in other study types; however, most of these studies cannot answer with sufficient certainty the relevant question as to whether a difference is caused by the intervention. Non-randomised studies always provide potentially biased results, even if only minor selection bias existed in the choice of participants. As a matter of principle, structural equality of groups cannot be assumed in these studies. The use of non-randomised studies as proof of the causality of an intervention therefore requires particular justification.

As a rule, the Institute therefore considers RCTs in the benefit assessment of drugs and only uses non-randomised intervention studies or observational studies in justified exceptional cases. It is argued therein that RCTs are usually possible and practically feasible in the assessment of drugs. Reasons for exception are, on the one hand, the non-feasibility of an RCT (e.g. if the therapist and/or patient have a strong preference for a specific therapy alternative) or, on the other, the fact that other study types may also provide sufficient certainty of results for the research question posed.

Meta-analyses of trials are often used to mathematically summarize the treatment effects of individual trials. Head-to-head comparisons of each treatment are not always available, so that indirect comparisons have to be included to ensure that all the pertinent treatments are considered. Appropriate methods should be used to derive these indirect estimates of treatment effects. The mixed treatment comparison (MTC) meta-analysis is preferably used in these cases. In MTC meta-analysis, direct and indirect evidence are combined to estimate treatment effects.

Given these data limitations or if cost data are unavailable from clinical trials, the use of alternative data sources for the estimation of resource use, e.g. from observational studies, may be considered. Due to their importance for model results, all data sources that are used to estimate resource consumption should be described and well justified.

In terms of the use of surrogate outcomes, as a rule, in the Institute's benefit assessments, surrogate outcomes are considered only as proof of a(n) (additional) benefit of an intervention if appropriate statistical methods applied beforehand showed that the effect of an intervention (with a comparable mechanism of action) on the patient-relevant outcome to be substituted was explained to a sufficient degree by the effect on the surrogate outcome. For this purpose, clear proof is normally required from intervention studies of a plausible, strong, consistent, and unidirectional association between the change in the surrogate outcome and the change in the patient-relevant outcome. The validity of a surrogate outcome is regarded as not proven if no relevant studies are available describing an association between the change in the surrogate outcome and the change in the corresponding patient-relevant outcome.

The Institute gives the highest evidence level to RCTs and systematic reviews of RCTs, at least within the framework of therapeutic studies. The next levels include non-randomised intervention studies, prospective observational studies, retrospective observational studies, non-experimental studies (case reports and case series) and, at the lowest evidence level, expert opinions not based on scientific rationale.

A benefit assessment on the basis of systematic reviews and HTA reports (both sometimes referred to as 'secondary literature' in the following text) can provide a resource-saving and reliable evidence base for recommendations to the Federal Joint Committee or the Ministry of Health, provided that specific preconditions have been fulfilled. The applicability of a benefit assessment prepared on the basis of systematic reviews and HTA reports depends on the availability of secondary literature that, in order to draw a clear conclusion:

- Is of sufficiently high quality and shows only minimum potential for bias
- Is directly relevant
- Includes high-quality primary studies

Health economic evaluations must allow for appropriate transferability of results to the German health care system, and must consider local conditions relating to epidemiology, health care resource availability, access to health provision, clinical practice, reimbursement of providers, and organizational structures.

In small populations (e.g. patients with rare diseases or special subgroups of patients with common diseases), it is argued that there is no convincing argument to deviate in principle from the hierarchy of evidence levels.

Comparators

Comparators that are relevant in a benefit assessment are usually the same as those to be used in a health economic evaluation. If, in contrast to the previous benefit assessment, the health economic evaluation produces benefit results on additional comparators in the relevant therapeutic area by means of adjusted indirect comparisons, these results can be included in the health economic evaluation depending on how robust they are.

Analysis of uncertainty / sensitivity analysis

The appraisal of both the qualitative and quantitative certainty of results, as well as the size of the effects observed (and their consistency), form the basis of the recommendations to be inferred and of their grading.

The impact of uncertainty on model results has to be investigated by means of sensitivity analyses. According to the Methods paper, there are three main types of sensitivity analysis in health economic evaluations: univariate and multivariate deterministic, and multivariate probabilistic (Monte Carlo).

It is not recommended to replace univariate sensitivity analyses with multivariate probabilistic sensitivity analyses. Instead, the latter should be carried out in addition to univariate analyses, if needed, so that the influence of individual important model parameters and assumptions remains visible. Finally, structural sensitivity analyses should be carried out to analyse the impact of a change in structural model assumptions.

All analyses undertaken to do so should be fully documented in terms of the ranges for the parameter values used and assumptions made. Models that go beyond the duration of underpinning RCTs are subject to greater uncertainty than models that are limited to the duration of an RCT. Accordingly, it is of increasing importance to adequately test and document the robustness of model results with longer modelling time spans.

Costs

The perspective of SHI insurants must be regularly included in the health economic evaluation. According to this perspective, the services covered by the SHI and the costs that SHI insurants have to pay themselves (out-of-pocket expenses) are reproduced. The cost estimate includes the following:

- direct reimbursable medical costs
- direct medical and non-medical costs that have to be borne by patients and their families (non-reimbursable costs)

Health economic evaluations carried out on behalf of IQWiG must also consider direct costs.

Indirect costs are not primarily considered. If loss of productivity is substantially affected by a new health technology, the corresponding costs may be evaluated separately. It should be noted that the costs are not to be listed on both the cost and benefit side. Loss of productivity due to mortality is included in the outcome on the benefit side. Loss of productivity due to incapacity for work is to be considered on the cost side as indirect costs.

If cost-offsets are taken into consideration, they should be investigated in comprehensive sensitivity analyses.

Time horizon

In principle, the health economic evaluation should cover the duration of the randomized controlled trials and, as a secondary scenario, be extended beyond this time period if this is relevant for the decision maker. The time horizon should appropriately reflect the natural course of a disease and be sufficiently long to capture all relevant benefit and cost considerations related to the health technology or programme

If required, IQWiG primarily carries out modelling for the time period for which evidence on benefit and harm from clinical studies exist. In a second step, health technologies can be modelled over longer periods of time.

Discounting

A discounting rate of 3% is stipulated. Sensitivity analyses have to be performed in order to examine the robustness of the results compared to the variation of this cost factor. These sensitivity analyses should be conducted for discount rates of 0 %, 5 %, 7 % and 10 %.

Budget impact analysis

Since neither the insurants' financial capacity nor their willingness to pay can be measured, the Institute is unable to make precise recommendations on the reasonableness for cost coverage. However, the Institute can assist this assessment by describing the possible future financial impact arising from this type of cost coverage. For this purpose IQWiG performs a budget impact analysis. The detailed budget impact analysis for determining a possible expenditure rate is on the basis of appropriate pricing/setting a maximum reimbursable price

Presentation of methods / results

A detailed technical report describing all the modelling steps from development of the initial influence diagram to final validation is required. In addition, a fully executable version of the model must be made available, along with a user manual.

References

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). (2008) *General methods*. Version 3.0. Cologne: IQWiG.

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). (2009) *general methods for the assessment of the relation of benefits to costs*. Version 1.0. Cologne: IQWiG. General Methods, Version 1.0, IQWiG, 2009

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). (2009) *Working Paper Modelling*. Version 1.0. Cologne: IQWiG.

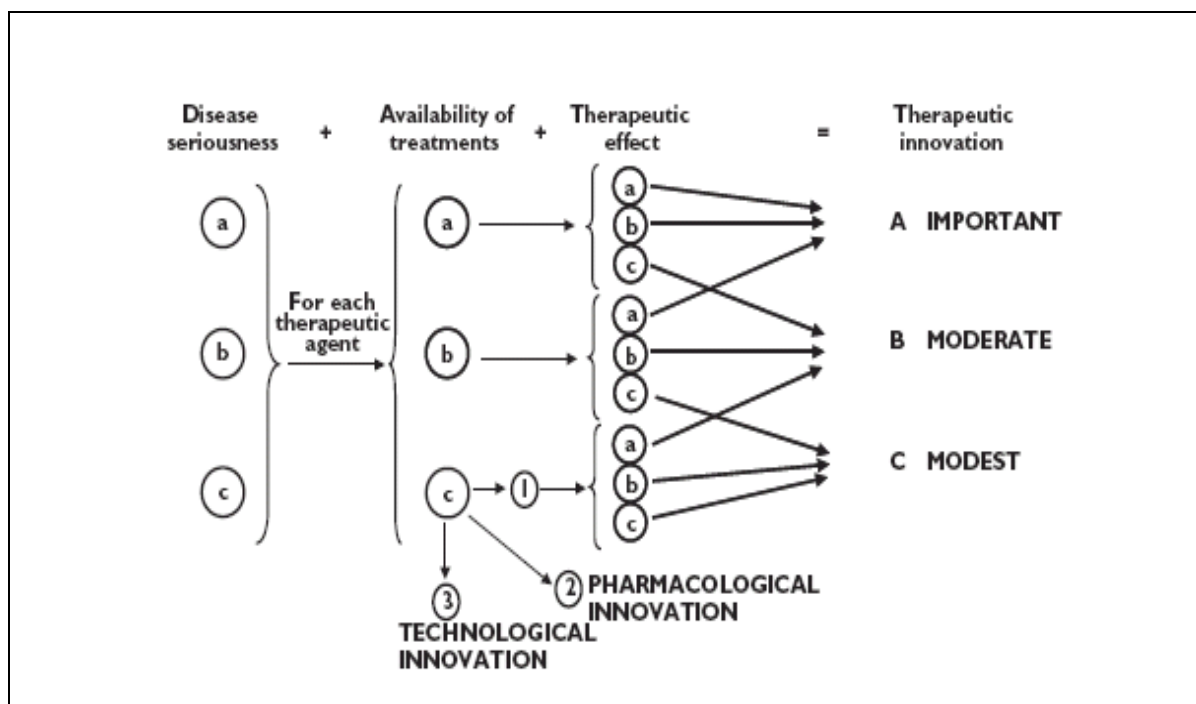
Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). (2009) *Working Paper Cost Estimation*. Version 1.0. Cologne: IQWiG.

Italy

Pricing and Reimbursement System: General Overview

The Italian approach to assessing the reimbursed price of innovative medicines continues to be essentially a pharmacological and formula led one, but a new model for classifying and scoring the therapeutic innovativeness of medicines was introduced by the Italian Medicines Agency (*Agenzia Italiana del Farmaco* [AIFA]) in 2006, based upon an expert paper published in 2005 (Motola, et al.)¹⁵. Figure 4 shows this algorithm.

Figure 4. Italian algorithm to assign the overall score for therapeutic innovation



The following factors are taken into account when assessing 'disease seriousness', 'availability of treatments' and 'therapeutic effect' respectively:

Disease seriousness:

- a: drugs for serious diseases e.g. neoplasms, Parkinson's disease, AIDS
- b: drugs for the treatment of risk factors for serious diseases e.g. hypertension, obesity and osteoporosis
- c: drugs for non-serious diseases e.g. allergic rhinitis

¹⁵ Motola, D., De Ponti, F., Rossi, P., Martini, N. and Montanaro, N (2005) Therapeutic innovation in the European Union: Analysis of the drugs approved by the EMEA between 1995 and 2003, *British Journal of Clinical Pharmacology*. 59, 475-488.

Availability of treatments

- a: drugs for diseases without recognised standard treatment
- b: drugs for diseases where subsets of patients are less responsive to marketed drugs and/or other medical interventions
- c: drugs for diseases responsive to marketed drugs or other medical interventions:
 - c1: more effective or safer or with better kinetics than existing drugs
 - c2: simple pharmacological innovation, i.e. a drug with a new mechanism of action but whose therapeutic role is comparable to existing products
 - c3: mere technological innovation, i.e. a new chemical or biotechnological product with a therapeutic role similar to already existing ones

Therapeutic effect

- a: major benefit on clinical end-points (e.g. increased survival rate and/or quality of life) or validated surrogate end-points
- b: partial benefit on the disease (on clinical or validated surrogate end-points) or limited evidence of a major benefit (inconclusive results)
- c: minor or temporary benefit on some aspects of the disease (e.g. partial symptomatic relief of a serious disease)

Thus, for example, if a medicine were considered by AIFA to be treating a serious disease (disease seriousness category 'a') but there is at least one other effective treatment for all relevant patient groups (availability of treatments category 'c') and the new medicine yields major clinical benefits (therapeutic effect category 'a') then overall it would be classed as 'B' – a 'moderate' therapeutic innovation (see Figure 4).

Three further points are worth mentioning. First, the introduction of the 'potential therapeutic innovation' concept within 'Pharmacological innovation' (c2) and 'Technological innovation' (c3), allows for more flexibility in handling uncertainty. For instance, a product might enjoy an early assessment of 'potential innovativeness', on the understanding that the company will submit a proposal, to be agreed with AIFA, in order to carry out additional researches and produce additional outcomes within two to three years.

Second, the 'three-year innovation window' was introduced, which implies that drugs launched in the same therapeutic area during the three years immediately following the authorisation of the leading product in that area will be accorded the same degree of innovation as the leading product.

Third, it is possible to change (increase or decrease) the degree of innovation attributed to a drug if new evidence is provided showing that its risk/benefit ratio has improved/ deteriorated.

Italy now only has three categories of reimbursed products: Category A and Category H (hospital only) which are reimbursed at 100% and category C reimbursed at 0% (i.e. the patient must pay the full price). Unlike in France, there is no correlation between the level of innovativeness and the level of patient co-payment.

Guidelines

No detailed guidelines were found.

The Netherlands

Pricing and Reimbursement System: General Overview

Medicines in The Netherlands are classified as List 1A, List 1B or List 2.

List 1A contains medicines subject to therapeutic reference pricing (i.e. price is set by reference to the prices of existing medicines that do the same job), and includes those medicines that are interchangeable. Medicines included in List 1B have a therapeutic added value compared to the standard or usual therapy and do not face direct price controls. The main criteria used currently to determine whether a medicine is included in List 1A or 1B are therapeutic value and reduction of side effects. The criteria underlying the assessment of therapeutic value include incremental efficacy, effectiveness, quality of life and safety, which in turn can include fewer/lesser side effects and greater tolerability.

There is a further category, List 2, which comprises conditionally reimbursed medicines. Criteria for admission to this category can be, for example, extremely high costs and high chance of improper use. Currently, only a limited number of products are included in List 2.

Examples of List 1A clusters of medicines include the statins and proton pump inhibitors, among others. Different compounds are clustered together with a single reference price, without regard to patent status, if all of the following criteria are met:

1. Same mechanism of action
2. Used for the same indication, based on actual use, not the official product labelling
3. Similar route of administration – for example, parenteral forms are grouped separately from oral forms of the same compound
4. Intended for the same age group
5. No significant differences in clinical effects. Differences are 'significant' if they affect physicians' choices

Once a medicine is deemed to be non-clusterable and hence is included in List 1B, the reimbursed price of this medicine cannot be higher than the manufacturer's maximum price, but it is then fully reimbursed.

Economic evaluation of health care programmes has become quite common in The Netherlands. In 1999, guidelines for pharmacoeconomic submissions were published by the 'College voor Zorgverzekeringen' (CVZ - Health Care Insurance Board) in co-operation with the industry. From 2002 to 2004, economic evaluation studies (pharmacoeconomic study and budget impact) were voluntary to seek a premium price and/or to have the product added to List 1B. From 2005, pharmacoeconomic studies and budget impact are formally required in the reimbursement qualification application of new, non-clusterable drugs to be included in List 1B and to seek a premium price. To coincide with this, the CVZ published new guidelines in 2005, which came in effect in April 2006.

As a result of these changes, the role of the CVZ has been strengthened, as it can now offer a specific yes/no recommendation with regards to whether a product should be reimbursed.

It is important to note that pharmacoeconomic studies are a 'second step' i.e. after relative therapeutic benefit is assessed. If a medicine is deemed a suitable candidate, in addition to its therapeutic value, its cost effectiveness and budget impact will be considered.

The manufacturers submit a dossier to substantiate their claim on therapeutic added value and cost effectiveness, compared to standard or usual treatment.

Guidelines

Analytical method – economic model

Outcomes of a cost-utility analysis (CUA) and/or cost effectiveness analysis (CEA), presented as an incremental costs effectiveness ratio (ICER) are recommended. The guidelines suggest that a CUA is appropriate to demonstrate differences in outcomes in terms of quality of life. Moreover, CUA is the preferred type of evaluation, with outcomes expressed in incremental costs per QALY gained. Cost-minimisation is not promoted, as the CVZ prefers at least a CEA.

There is no cost-effectiveness threshold.

Health outcomes / QoL measure

QoL may be described using generic and disorder specific questionnaires. As a societal perspective should be used, a representative random sample of the population is the most suitable source for the utility instrument. QoL results must be reported separately and QALYs are recommended in the primary analysis.

Perspective

A societal perspective should be used for the economic model.

Evidence considered

Pharmacoeconomic studies should report on effectiveness, and that data should be collected ideally under realistic conditions. Modelling is allowed to translate data from efficacy studies, but all assumptions must be explicitly stated and be subject to sensitivity analysis.

Expert panels are allowed to obtain missing information, although the members of this panel should be described.

Comparator

The standard or usual treatment (which may be a pharmaceutical product or procedure) should be used as the comparator. The standard treatment is defined as the treatment in daily practice that has been identified as the 'first choice for which effectiveness has been proven'.

Choice of patient group

Pharmacoeconomic evaluations should be carried out on the entire study population as well as any subgroups identified in the protocol, on the basis of any possible differences in effectiveness, costs or other variables.

Analysis of uncertainty / sensitivity analysis

All underlying assumptions and methods should be explained carefully. The uncertainty of deterministic variables should be assessed using univariate sensitivity analysis. In a modelling study, probabilistic sensitivity analyses are required to determine the uncertainty of the variables used.

Costs

Direct healthcare and broader non-health care direct costs are to be included. Any healthcare indirect costs resulting from illnesses not related to the intervention should be excluded.

Time horizon

The time horizon used in the modelling must enable valid and reliable statements to be made about the product.

Discounting

Discount rates to be used are 4% for costs and 1.5% for outcomes. But the guidelines state that these rates should be subject to sensitivity analysis.

Budget impact analysis

The CVZ carries out its own budget impact analysis to complement any done by the industry.

References

IMG Health. (2007) Unlocking the potential of evidence-based medicine in a cost-contained market. *IMS Pharmacoeconomics Report*.

Scotland

Pricing and Reimbursement System: General Overview

Scotland operates its own healthcare system under the political and administrative direction of the Scottish Parliament, although there continue to be complex funding relationships with the UK government in London. A notable consequence of this is that the Scottish NHS operates its own HTA Agency which performs a role equivalent to NICE, called the Scottish Medicines Consortium (SMC). It assesses added value and takes a view of affordability and access to reimbursement, completely independent from NICE. (www.scottishmedicines.org.uk)

The SMC operates to much the same pharmacoeconomic criteria as NICE in which the cost per QALY plays a central role, but has less bureaucratic processes and is inclined to take faster and more judgemental views as to added value based upon a clinical and economic perspective. According to its website, the SMC aim is to issue advice to NH Scotland on all newly licenced medicines within 12 weeks of products being made available.

Medicines can only be approved if the manufacturer makes a submission to the SMC. If the manufacturer does not make a submission then SMC will advise that use of the product cannot be recommended - thus the onus is on the manufacturer to proof 'value for money'. The SMC does not carry out its own analysis but rather provides in its guidance a critique of the submission. This has led to faster access in Scotland than in England for some new medicines.

There are two types of submission, Full or Abbreviated:

- Full submissions are necessary for a new chemical entity (even if the number of potential patients is small and the expected cost is low); additional indications approved by the MHRA/EMA (again, even if the number of potential patients is small and the expected cost is low); new formulations (e.g. slow release, liquid) at more than pro rata cost per treatment; and new biosimilar medicines
- Abbreviated submission are used for follow-on biosimilar medicine (when the SMC has already accepted the existing biosimilar for use for the same indication) at pro rata cost or less per treatment; new formulations (e.g. slow release, liquid) at pro rata cost or less per treatment – clinical equivalence or superiority must be capable of being demonstrated briefly, in simple terms; and a medicine indication extended for use in children or adolescents, where product has previously been accepted by SMC for use in adults

The New Drugs Committee (NDC), a subgroup of SMC, undertakes a rapid assessment of the evidence provided by the drug manufacturer and prepares draft advice for the SMC to consider about the costs and benefits of using the medicine.

When the SMC has accepted a new medicine, NHS Boards are expected to make it (or its equivalent) available. When the SMC offers a 'Not Recommended' Advice, NHS Boards are not expected to make it routinely available. However, medicines 'not recommended' by SMC, including those medicines 'not recommended' due to a non-submission, can be made available under certain circumstances

through individual patient treatment requests. Where medicines have been ‘not recommended’ by SMC, it remains open to the pharmaceutical company to submit or re-submit evidence about the medicine to the SMC at any time.

Overall, standard decision rules should be followed in combining costs and QALYs. These should reflect any situation where dominance or extended dominance exists. Incremental cost-effectiveness ratios (ICERs) reported must be the ratio of expected cost to expected QALY.

In assessing the relative clinical- and cost-effectiveness of new medicines, the Scottish Medicines Consortium (SMC) requires a robust clinical and economic case to be made and for the medicine to demonstrate value for money. In some specific situations SMC may exercise greater flexibility in its decision making to allow consideration of additional factors. These may allow SMC to accept either more uncertainty in the health economic case or a higher cost per Quality Adjusted Life Year (QALY).

Guidelines

The SMC has recently published a revised set of guidelines (February 2010). It states that the SMC requires manufacturers to submit economic evaluations consistent with its Guidance. Compliance with the SMC Guide is mandatory. The SMC has not specified a reference case which must be adopted as a base case. This reflects the emphasis SMC places on receiving submissions as close as possible to the time of launch: such timing may preclude manufacturers from presenting all the data required for a pre-specified reference case.

However, to assist manufactures, the SMC has judged that the reference case set out by NICE (see above) is appropriate for use in a submission to the SMC. The key elements of the analysis in the NICE reference case are summarised in Table 2.

Table 2.

Element of Health Medicine Assessment	Reference Case
Comparator	Alternative therapies routinely used in the NHS in Scotland
Perspective on costs	NHS in Scotland and social work
Perspective on outcomes	All health effects on individuals
Type of economic evaluation	Cost-utility analysis
Synthesis of evidence on outcomes	Based on a systematic review
Measure of health benefits	Quality-adjusted life years (QALYs)
Description of health states for calculation of QALYs	Health states described using a standardised and validated generic instrument

Method of preference elicitation
for health state

Choice-based method, for example, time trade-off

Analytical method

In general SMC's preferred format of economic evaluation is cost utility analysis, with health effects expressed in terms of quality-adjusted life-years (QALYs). In some circumstances, more simple economic evaluations such as cost-minimisation analysis may be appropriate if the proposed medicine is demonstrated by studies to be therapeutically equivalent to the relevant comparator(s), as assessed using an adequately designed and powered non-inferiority or equivalence or superiority study. Alternative approaches can be considered in those circumstances in which the QALY may not be the most appropriate outcome measure.

However, the guidelines state that manufacturers are urged to think carefully before deciding not to use QALYs as the SMC regards this methodology the most appropriate to make comparisons of value across health care interventions. If manufacturers present other methods (e.g. willingness to pay studies or a discrete choice experiment) these must be fully described and the uncertainty in results fully explored.

Cost-consequence analysis is not useful to the SMC as the trade-offs between different dimensions of benefit are not made clear.

Overall if the manufacturer does not satisfy the SMC on these basic design points then no amount of complex modelling or sophisticated sensitivity analysis will compensate for this. However, even if the economic case is well-conducted and robust, the economic evaluation needs to demonstrate that the product offers reasonable value for money to NHS Scotland compared to other uses of scarce NHS resources.

Patient access schemes will be considered by NHS Scotland to facilitate access by patients to medicines (i) that are not, or might not be, in the first instance found to be cost-effective by SMC or (ii) where a patient access scheme has been accepted in the context of a NICE Multiple Technology Appraisal.

Health outcomes/QoL measure

To promote consistency across appraisals, SMC has a preference for generic measures of outcome, such as the quality-adjusted life-year (QALY), over disease-specific measures of outcome, suggesting that the manufacturer should give serious consideration to a cost-utility analysis (CUA). This does not mean that CUA is required every time. Some examples of where it could be argued that it is unnecessary include when the QALY does not capture the main benefit of the medicine (e.g. contraception) or when utility values appear to lack sensitivity in circumstances where other measures suggest health improvements or disease reductions. SMC would need to be assured that the changes on the non-QALY measures are valued by patients.

The guidelines state that there may be circumstances where a simpler approach than CUA can be used but these should be considered with care by manufacturers. One approach is a cost impact study concluding there are net economic costs savings, in addition to the clinical evidence suggesting the medicine is at least as good as the alternative (e.g. a 'me-too' drug). This approach clearly has merits but it is not without risk. The same could be said of a cost-minimisation analysis (CMA). This might be used where a clinical trial shows no difference in the main outcome. However, there may be objective organisational or patient issues that lead SMC to conclude that there are other aspects of benefit beyond the primary outcome selected; again, this might make a simple economic analysis seem inadequate.

The guidelines also state that it would be advantageous if the submission includes results that directly measure improvements in the patient's quality of life rather than in a proxy measure. SMC places great weight on the patient's perspective on health gain.

If utility data from generic validated instruments is not available, the SMC will, in general, accept utilities from three other sources:

- Utilities mapped from a disease specific quality-of-life measure included in a clinical study - the SMC will want to see well designed and explicit methods of mapping from the disease-specific measure to a generic measure and from there to utilities
- Specific surveys for direct measurement of utilities for appropriate disease/condition health states. This should use time trade off (TTO) or standard gamble (SG) methods of utility elicitation. SMC will accept values from either public members or patients and places more store by the perceived validity of the utility values when put in the context of utilities for other health states. The SMC need a description of the vignettes of health states used for the valuation and a clear explanation of how the health states were derived.
- Values taken from previous studies reported in published literature. However, the submission must report all of the utility values reported in the literature and the literature selection process, in order that the SMC can see that the manufacturer has not been selective. The submission must also show that the health state valued in the literature reflects the health states in the submitted economic evaluation. For example, if the new medicine is for advanced prostate cancer it is not sufficient to use literature values that are reported for the state 'prostate cancer' with no further description.

Perspective

The perspective on outcomes should be all direct health effects whether for patients or, where relevant, other individuals (principally carers). The perspective adopted on costs should be that of the NHS in Scotland and social work (referred to as Personal Social Services (PSS) in England). If the inclusion of a wider set of costs or outcomes is expected to influence the results significantly, these should be reported in a sensitivity analysis.

Evidence considered

The guidelines distinguish between four key dimensions: efficacy, clinical effectiveness, comparative efficacy and clinical effectiveness.

For efficacy, companies need to provide details of studies, which provide evidence of the clinical benefits with the drug in the indication(s) under review relative to active comparator(s) used in clinical practice. The guidelines state that the most relevant are active-controlled trials. However, if active-controlled trials are not available, details of placebo-controlled or uncontrolled trials should be included. Placebo-controlled and uncontrolled trials can also be included if they provide evidence of relevant clinical benefits not demonstrated in active-controlled trials. Moreover, details of ongoing studies or updated analyses of trials described should be acknowledged, which would provide additional evidence within the next 6 to 12 months for the drug in the indication(s) under review.

In terms of clinical effectiveness, companies need to state whether data on the clinical benefits and adverse effects with the drug in the indication(s) under review relative to relevant comparator(s) were available from active-controlled trials.

Indirect comparison can be used in the economic model, especially if no head to head evidence is available. If this is the case, companies need to provide details of the search strategies or rationale for identification of data sources and on data sources themselves used in the indirect comparison to provide evidence of clinical benefits and adverse effects. Companies also need to provide details of any relevant differences between the data sources providing evidence of clinical benefits and adverse effects with the drug in the indication(s) under review and those providing evidence for indirect comparator(s). These would include, but not be limited to, differences in terms of (a) patient populations; (b) drug treatments; (c) methodology; (d) study limitations; and (e) result.

For evidence on comparative efficacy, companies should include details of randomised controlled trials (RCTs), meta-analyses and other studies, that provide evidence of the clinical benefits of the drug in its licensed dose within the indication(s) under review relative to active comparator(s) used in routine clinical practice. The most relevant are active-controlled studies. However, if active-controlled studies are not available, details of placebo-controlled or uncontrolled studies, that provide evidence of the clinical benefits of the drug in its licensed dose within the indication(s) under review, should be included. Placebo-controlled and uncontrolled studies can also be included if they provide evidence of relevant clinical benefits not demonstrated in active-controlled studies.

For clinical effectiveness, companies should provide details of whether studies have directly measured health outcomes such as mortality, survival, incidence of disease, morbidity, functional performance, quality of life or whether surrogate markers have been measured e.g. reduction in blood pressure. They should also provide details of any association between surrogate markers and health benefits or disadvantages to patients.

In general, companies should also provide details of differences between the patient populations included in the studies that provided evidence of clinical benefits and adverse effects compared to

the Scottish population likely to receive the drug in clinical practice. Examples of this include, but are not limited to, the following:

- Differences in baseline demographics, such as age, performance status, previous treatments, severity of disease
- Differences in clinical management, such as the dose schedule of comparator(s) or permitted/disallowed concomitant drugs, monitoring or assessment frequency

Additional evidence accepted includes that coming from:

- **Systematic Reviews:** Databases searched and literature searching strategies should be reported. There should be a clear rationale for selecting specific studies from those identified
- **Meta-analysis:** Synthesis of outcome data through meta-analysis is appropriate provided there is sufficient, relevant and valid data that uses comparable measures of outcome
- **Role of Expert Opinion:** Where data from studies are insufficient to provide values for relevant variables, and such values can be obtained from expert opinion, then SMC will consider this as a valid source of evidence. The impact of this evidence will be greater if the submission is transparent on the process used, for example on the selection criteria used to approach potential experts and the range of values provided by the experts. Variables elicited from expert opinion should be tested in the sensitivity analysis.

The SMC requires some reassurance that resource use data used are broadly representative of Scottish patient pathways and clinical practice. Data from elsewhere in the UK are acceptable. Resource use data from other countries or estimated by a panel of experts are regarded with some suspicion and should be avoided if possible, or at least validated for the Scottish setting and included in a rigorous sensitivity analysis.

Comparator

The comparator the SMC is interested in is the one that will most likely be replaced if the medicine under consideration is accepted by the SMC for use in Scotland. This may be different to the comparator in the clinical studies programme for the medicine; if so, the manufacturer must carry out an indirect comparison. The service replaced might also involve no active treatment of the condition.

The guidelines also state that there are frequently several potential comparator medicines as, for example, practice is not necessarily consistent across Scotland or the UK and between the UK and elsewhere. All relevant comparators must be identified, although a full comparison will not always be appropriate for every one of these comparators.

Choice of patient group

We did not find any detailed information about this issue.

Analysis of uncertainty / sensitivity analysis

The guidelines state that it is important for the SMC to understand the uncertainty associated with the clinical and cost-effectiveness information. This includes uncertainty about the clinical and cost-effectiveness estimates, choice of studies to include in a meta-analysis, and the structural assumptions made in a model. A rigorous sensitivity analysis may demonstrate the robustness of the base case result. Appropriate use should be made of one-way, scenario, threshold and probabilistic sensitivity analysis.

The SMC require the manufacturer to demonstrate, through the use of sensitivity analyses, of:

- The robustness of the ICERs
- Under which circumstances the ICER exceeds £20,000 and £30,000.

NICE describes an approach to synthesise data into a point estimate of treatment effect plus the variance around that estimate. This approach is valid but the SMC will also accept an economic evaluation based on a single clinical study provided,

1. The patients recruited to the study are broadly representative of a Scottish or UK population
2. The study is a 'head-to-head' with the relevant comparator, and
3. It is demonstrated by the manufacturer that this study does not have notably different results to the studies that are not used. In other words, the SMC requires consistency of clinical effects.

Costs

Where possible, costs should relate to resources that are under the control of the NHS in Scotland and social work (equivalent to PSS in England) where differential effects on costs between the medicines under comparison are possible. These resources should be valued using costs relevant to the NHS in Scotland and social work. Where this is not possible, UK data may be adapted based on Scottish population statistics.

Costs will be for continuous treatment for one year (52 weeks) with some exceptions, e.g. for treatment given in courses and for seasonal therapy. In those cases, the treatment period used for costing will be based on the assessor's clinical judgement and/or on expert advice. The costing for each drug will be based on the product, formulation, strength and pack size which gives the lowest cost, provided that it represents a realistic choice for use in practice. This may be a generic product and, in general, the cost of more expensive products will not be given. However there may be exceptions e.g. for clinically relevant alternative formulations.

It is current policy for the cost comparisons to be based on prices used for NHS reimbursement (which for proprietary products is the list price), and to obtain list prices for those products not covered by this e.g. hospital-only products. While SMC recognises that this may not reflect the costs being paid in practice for some products used in hospitals, it provides a more even comparison between products.

Assessment of cost effectiveness and resource implications may take account of contract or discounted prices as appropriate.

Time horizon

It is necessary that clinical and cost-effectiveness are considered over an appropriate time horizon relevant to Scottish practice and patients.

Discounting

The annual discount rate recommended for both costs and benefits is 3.5%. When results are potentially sensitive to the discount rate used, sensitivity analysis should vary the rate between 0% and 6%.

Budget Impact

An SMC submission always requires a budget impact assessment.

Liaison between SMC and manufacturer

There is no 'face-to-face' contact between manufacturers and reviewers within the process. However, reviewers have the opportunity to contact manufacturers for further information or clarification. According to the guideline, experience to date has shown the value in this contact. This mainly takes place prior to the NDC meeting and is usually carried out via the secretariat.

References

Scottish Medicines Consortium. (2010) *Guidance to manufacturer on submission criteria*. Glasgow: Scottish Medicines Consortium. [Accessed 26 July 2010].

Healthcare Policy and Strategy Directorate, Quality Division. (2010) *Introduction and availability of newly licensed medicines in the NHS in Scotland*. Edinburgh: The Scottish Government. 17 May.

Scottish Medicines Consortium. (2006) *Cost of relevant comparators' in the Detailed Advice Document from the Scottish Medicines Consortium: Frequently Asked Questions*. July. Glasgow: Scottish Medicines Consortium.

Scottish Medicines Consortium. (2010) *Guidance to Manufacturers: Question and answer document on economic submissions to the SMC*. June. Glasgow: Scottish Medicines Consortium.

Spain

Pricing and Reimbursement System: General Overview

The Spanish Ministry of Health sets the maximum ex-factory price of medicines product by product after a negotiation with the manufacturer. Formally, the price is set according to a cost plus system which states that a drug's price must cover all costs (raw materials, production, administration, promotion, management, research and development, general costs) plus a profit margin. The price

must be consistent with the therapeutic utility of the product and with the price of alternative treatments.

The following factors determine whether or not a medicine is reimbursed by the Spanish Ministry of Health (and is thus included in the positive list¹⁶):

1. The seriousness, duration and consequences of the various disorders
2. The needs of certain groups
3. The medicine's therapeutic and social utility
4. The limits of public expenditure allocated to pharmaceutical benefits
5. The existence of medicines or other product alternatives for the same conditions
6. The medicine's degree of innovation

In addition, another criterion included in the same Article of the Law (but not as a listed point) is the price or cost of the comparable medicines available in the market.

It is worth noting that the new Medicines Act, approved on July 26, 2006 (which replaced the 1990 Law) incorporates for the first time the medicine's 'degree of innovation' as a factor to determine a medicine's reimbursement status. However, in the new Act, there is no discussion as to what factors will be considered when assessing a product's degree of innovation.

The 2006 Medicines Law stated that a new advisory Committee to Assess the Therapeutic Value of Medicines was to be established with nominated regional experts. However, to date, this Committee has not been set up officially.

A consortium led by Andalusia, Catalonia and the Basque country has created regional HTA agencies (giving rise to the so-called 'Mixed Regional Committee') which classify medicines into the following categories:

- 0) Insufficient information to classify
- 1) Does not represent a therapeutic advance
- 2) A product for use in restricted indications or patient groups
- 3) A modest therapeutic advance
- 4) A major therapeutic advance

Note, however, that these Committees can only publish recommendations, as the pricing and reimbursement decision is taken at the national level.

These organisations also carry out limited 'economic analyses' in the form of comparing daily treatment costs. Overall the type of analysis is very restricted. A retrospective review over the period 1994-2006 by the three main agencies of 245 evaluations shows that only 10 out of these 245

¹⁶ The positive list includes all the publicly reimbursed medicines in Spain.

were rated as major advances (4%) whereas 55% were classified as not offering any therapeutic advance (Jönsson, et al., 2007)¹⁷.

Guidelines

No detailed guidelines were found.

Sweden

Pricing and Reimbursement System: General Overview

In Sweden HTA is the prime basis for admission to the reimbursement system for new products. Sweden administers market regulation for drugs through a single agency, formerly known as LFN, and now renamed Dental and Pharmaceutical Benefits Agency with the acronym TLV. The law sets forth the criteria which must be fulfilled if a medicine should be reimbursed. Those criteria consist mainly of three principles:

1. The human value principle; care should be given with respect for the equality of all human beings
2. The solidarity principle; those in greatest need take precedence in medical care
3. The cost-effectiveness principle; the cost for using a medicine should be reasonable and fair from a medical, humanitarian and social-economic perspective

The main rule granted by TLV can either be to recommend a general subsidy or to reject a health technology. However, TLV can recommend restriction, an example of which may be that the same medicinal product is used by two or more completely different indications and that the treatment leads to different levels of patient benefit and/or demonstrates totally different cost-effectiveness.

In October 2003, the LFN/TLV started a five-year cost-effectiveness review to see whether some 2,000 products still merit reimbursement – but in practice it has taken much longer than five years and only a handful of reviews have been published to date. This cost-effectiveness review is part of the reimbursement regulation introduced in October 2002. Products are reviewed according to the disease group they belong to, and are judged on medical effect as well as cost-effectiveness. It was announced that first groups to be reviewed in 2004 were migraine, gastro-intestinals, anti-hypertensives, anti-asthmatics and antitussives, followed in 2005 by antidepressants, cholesterol lowerers, analgesics and anti-inflammatories, antidiabetics, and incontinence and prostate therapies. Deadlines for these reviews have not been met. Indeed, as of July 2010, only seven out of the 49 therapy reviews have been completed¹⁸.

¹⁷ Jönsson, B., Staginuss, U. and Wilking, N. (2007) Acceso de los pacientes a los farmacos contra el cancer en Espana, *Revista Espanola de Economia de la Salud*, Monografico, 6(3), 175-183.

¹⁸ The seven completed therapy reviews are for: Migraine; Diseases caused by excess stomach acid; Asthma, COPD and coughs; Hypertension; Depression; Lipid disorders; and Diabetes. These reports can be found at: <http://www.tlv.se/in-english/reimbursement-review/avslutade-genomgangar/> (accessed 28 July 2010).

Guidelines

Analytical method

The application must include a health economics analysis. Cost-effectiveness analysis is recommended, with quality-adjusted life years (QALY's) as the measure of effect. In treatments that mostly affect survival, both QALY's and gained life years should be shown.

If it is difficult to use QALY's (e.g. with heavy pain over a short time in connection with treatment), then a cost-benefit analysis with the willingness to pay may be used as a measure of effect.

If there is supporting evidence that the drug to which the application refers has the same health effect as the best comparable treatment, a cost comparison may suffice.

If the medicine has the same effect as the comparison alternative, cost minimisation analysis is sufficient.

For products with a very low expected sales value, it may be sufficient to roughly specify expenses and the effect of the actual product compared with the relevant comparison alternative.

Analyses based on good empirical data carry weight. However, as stated in the guidelines, it is sometimes necessary to use modelling so the health economics analysis covers the relevant timeframe. Modelling is sometimes useful for achieving better external validity in clinical trials (adjusting for differences between clinical trials and clinical practice), or for adjusting clinical trials conducted in another country to Swedish conditions, e.g. it can be appropriate to use a health economic model to combine information regarding a therapy's effect from international randomised clinical trials with specific Swedish information about treatment practice, costs and characteristics of the patient population. Models should, as far as is possible, be validated internally and externally.

Health outcomes/QoL measure

QALY-weightings should be based on methods such as the Standard Gamble (SG) or Time-Trade-Off (TTO) methods. In a second instance, QALY-weightings should be based on the rating scale method. QALY-weightings can be based either on direct measurements with the above-mentioned methods or indirect measurements (where a health classification system such as EQ-5D is linked to QALY weightings). QALY weightings based on appraisals of persons in the health condition in question are preferred before weightings calculated from an average of a population estimating a condition depicted for it (e.g. the 'social tariff' from EQ-5D). Using weightings for current health conditions collected from previous studies may be a solution.

Perspective

Analyses should take a social economic perspective. All relevant costs and revenues for treatment and ill health, irrespective of the payee (county council, local authority, state, patient, relation) should be considered. Moreover, the information must describe the situation in Sweden.

Evidence considered

If existing randomised clinical trials do not offer a relevant treatment alternative for Swedish conditions, the analysis should be supplemented by a model calculation. The calculations carried out should be shown so that the assumptions and procedure are evident.

If the results from randomised clinical studies only contain partial amounts of the patient population to which the application refers, modelling should be undertaken to illustrate cost-effectiveness in the remaining patient groups. An estimation of the number of persons in each patient group in Sweden should be attached.

If surrogate end-points are used, the submission should also include modelling from these end-points (e.g. survival) to illustrate the effects on mortality and morbidity, i.e. QALY's gained.

The guidelines include the evidence that manufacturers need to submit in their application to TLV. Among others, companies need to include a brief description of all clinical studies and their results (pivotal phase 2 and all phase 3 studies) plus references to reported studies, as well as all relevant studies in their entirety that may clarify the clinical effect and negative effects of the medicine and that are relevant to the health economics analysis. Account of ongoing and/or planned studies also needs to be included. The guidelines argue that the best documentation includes directly comparative studies with the most relevant comparison alternative. If no direct comparative studies have been made, it is possible to use indirect comparisons of systematic overviews, for example.

A health economic study that has been peer reviewed and published in an international scientific journal has undergone a form of quality control. If the study is unpublished, greater demands are placed on the possibility for quality control and transparency.

Comparator

The costs and health effects of using the drug in question should be compared with the most appropriate alternative treatment in Sweden (e.g. the most used). This could be drug treatment, another treatment or no treatment at all. In making calculations, the reference point should be practice applicable in Swedish medical treatment.

Choice of patient group

Separate calculations should be made for different patient groups where the treatment is expected to have different cost-effectiveness (e.g. separately for men and women in different ages and with differing degrees of severity for the illness/symptom or with different risk levels). If the results from randomised clinical studies only contain partial amounts of the patient population to which the application refers, modelling should be undertaken to illustrate cost-effectiveness in the remaining patient groups. An estimation of the number of persons in each patient group in Sweden should be attached.

Analysis of uncertainty / sensitivity analysis

The sensitivity analysis of central assumptions and parameters is an important stage in health economic analysis.

Costs

All relevant costs associated with treatment and illness should be identified, quantified and evaluated. The production loss for treatment and sickness should also be included (estimated using the human capital method). Unit costs and quantities should be presented separately as far as is possible so that a distinction can be made between price and quantity. It should be clear what year prices represent. Apoteket's Sales Price (AUP) for medicine must be used. If the treatment affects survival, then the costs for increased survival – total consumption less total production during gained life years – should be included.

Time horizon

The timeframe for the study should cover the period when the main health effects and costs arise.

Discounting

Both costs and health effects should be discounted by 3 per cent. In the sensitivity analysis, the calculation should also be carried out using 0 and 5 per cent, as well as a calculation where costs are discounted by 3 per cent and health effects by 0 per cent.

Budget impact analysis

We could not find a reference for this issue.

Presentation of methods and result

Methods, assumptions made and detailed data should be shown so clearly that the different steps in the analysis are easily followed. Cost-effectiveness ratios should be calculated based on the differences in costs and effects (QALY's) that exist between treatment alternatives (incremental analysis).

TLV / company liaison

The government office usually has contact with the company while they are preparing the case. The purpose of these contacts is to clarify the circumstances in the case and to obtain further information from the company, as well as to give the company certain guidelines concerning the relevant questions. Sometimes, it is appropriate for the office to meet up with the company to examine the questions that have arisen. Companies are invited for deliberation when the proposed decision that has been made is a rejection or a subsidy that is linked with conditions or restrictions. Guidelines state that it is not appropriate to submit new documentation on the case at the time of the deliberation.

References

Pharmaceutical Benefits Board (LFN). (2003) *General guidelines for economic evaluations from the Pharmaceutical Benefits Board*. Stockholm: LFN.

Pharmaceutical Benefits Board (LFN). (2008) *Guidelines for Companies*.

Appendix 2. Search Strategies for the Literature Review

Medline

EFPIA_medline 10/05/10

Ovid MEDLINE(R) <1996 to April Week 4 2010>

1. effectiveness.ti,ab.
2. generalisab\$.ti,ab.
3. generalizab\$.ti,ab.
4. (external adj validity).ti,ab.
5. transferability.ti,ab.
6. transferable.ti,ab.
7. transferrable.ti,ab.
8. or/2-7
9. europe.ti,ab.
10. european.ti,ab.
11. 9 or 10
12. 1 and 8 and 11
13. review.pt.
14. 1 and 8 and 13
15. 12 or 14
16. limit 15 to (english language and humans and yr='2000 -Current')

Embase:

Same as medline; 10/05/10

EMBASE <1996 to 2010 Week 18>

Econlit:

- 1 effectiveness.ti,ab.
- 2 generalisab\$.ti,ab.
- 3 generalizab\$.ti,ab.
- 4 (external adj validity).ti,ab.
- 5 transferability.ti,ab.
- 6 transferable.ti,ab.
- 7 transferrable.ti,ab.
- 8 or/2-7
- 9 europe.ti,ab.
- 10 european.ti,ab.
- 11 9 or 10
- 12 1 and 8 and 11
- 13 1 and 8
- 14 review.tw.
- 15 13 and 14
- 16 12 or 15
- 17 limit 16 to yr='2000 -Current'
- 18 limit 17 to english
- 19 from 18 keep 1-9

HMIC:

HMIC Health Management Information Consortium <January 2010>

10/05/10

- 1 effectiveness.ti,ab.
- 2 generalisab\$.ti,ab.
- 3 generalizab\$.ti,ab.
- 4 (external adj validity).ti,ab.
- 5 transferability.ti,ab.
- 6 transferable.ti,ab.
- 7 transferrable.ti,ab.
- 8 or/2-7
- 9 europe.ti,ab.
- 10 european.ti,ab.
- 11 9 or 10
- 12 1 and 8 and 11
- 13 1 and 8
- 14 review.ab,pt,sh,ti.
- 15 13 and 14
- 16 12 or 15
- 17 limit 16 to yr='2000 -Current'
- 18 from 17 keep 1-42

Appendix 3. Key Points from Eichler, et al. (2010)

1. The HLPF report 'creates a space for RE (relative efficacy) assessments somewhere between regulatory and health-economic assessments, and calls for development of a common methodology that would allow RE estimates to be more broadly applicable than to merely one healthcare environment (as is full health-economic analysis). Is silent as to who should conduct RE assessment in the EU context . . . (and) the EMA has been given a political mandate to interact with HTA bodies.'
2. Relative Efficacy requires a) head-to head RCTs or b) post-launch pragmatic RCTs or c) indirect comparison using RCT data or d) observational studies. There are issues as to which comparator, if companies had to test against 'several active comparators in head-to-head RCTs as a basis for marketing authorisation ...this would raise the bar so high as to effectively shut down the drug development endeavour.'
3. In reality EMA can ask for active comparator RCTs, but does not all that often.
4. Companies are planning more active comparator Phase III trials. 'Increased harmonisation of study protocols within a given therapeutic area may facilitate pre-planned meta-analysis and indirect comparisons for RE (relative efficacy) assessment by payers'.
5. EMA Risk Management Plans (US REMS) may lead to 'Many post-marketing studies conducted under a regulatory REMS/RMP may lend themselves to an integrated evaluation of benefits and risks under real-world conditions.' They link this to US CED with a requirement for real world studies. 'We anticipate that REMS/RMPs will evolve over time into effectiveness- and, ultimately, relative (or comparative)- effectiveness-evaluation-plans. Joint input from regulatory agencies and payers into the design of these post-marketing studies would satisfy the regulators' needs ... and inform repeat cycles of coverage decisions of new drugs by payers.'
6. 'As is the case for the current REMS/RMPs, future relative effectiveness plans are expected to encompass the full spectrum of study methodologies, including (adaptive phase III/IV~ RCTs, PCTs, indirect comparisons/meta-analyses. This holistic approach will prove challenging, as it involves combining and weighting of information from very disparate data sources, but also because it will raise new questions of funding of post-marketing studies and ownership of study protocols and data.'
7. '..we anticipate that the current political momentum in the EU and US will create a new space for RE assessment/CER, bringing about a two-stage process between the licensing and reimbursing decisions: first, an assessment of RE/CER, with a focus exclusively on the medical consequences of competing treatment options, followed by a second analysis of context-specific economic consequences undertaken by individual payers... in the EU, healthcare policy is within the remit of individual member states, and the healthcare environments and purchasing power of payers are still quite diverse across member states. Hence, EU-wide harmonisation of reimbursement decisions is not likely to be readily or quickly achieved, but there is growing realisation that RE assessment is more an issue of scientific methodology than policy making, and could potentially be harmonised.'
8. 'It is difficult to speculate *who* will take a leadership role in RE/CER assessment. Time will tell whether, in the different jurisdictions, this task will be taken on by a dedicated entity (e.g.

institute or agency for RE/CER), by regulatory agencies, by existing HTA bodies or payers, or a combination of these. ‘

9. ‘...we anticipate that several paradigm changes will be required from industry, regulators and payers to address the RE challenge:
- Manufacturers will be required to prove that their products are not interchangeable with cheaper compounds on the market. A business model based on the RE paradigm is expected to trigger a reallocation of R&D resources away from me-too drugs, towards drug development programs which aim at superiority claims, often for smaller treatable populations, but with a predefined margin of improvement in benefit-risk.
 - RCTs have become overly complex, time consuming and inefficient. For example, it has been estimated that site-monitoring absorbs 25-30% of Phase III trial cost. As RE/CER requirements will be incorporated into clinical development, trials will inevitably grow bigger in size (box 2) and cost. To ensure that RCTs are still feasible, inefficiencies need to be remedied. Very timely the FDA has initiated a public-private partnership, the Clinical Trial Transformation Initiative (CTTI), which is currently exploring ways to enhance operational efficiency of RCTs.
 - The current, rigid model of confirmatory trials is not well suited to satisfy regulators’ and payers’ information needs. We have illustrated potential advantages of adaptive phase III/IV trials (Box 2). Stakeholders will need to agree on such or similar models of ‘learning adaptive’ RCTs to enable RE assessment without raising the evidence barrier to unrealistic levels.
 - All stakeholders, including academic groups will need to agree on and apply common research standards for non-interventional approaches to RE assessment, i.e. indirect comparisons, network meta-analysis, and observational studies. This may encompass some form of pre-registration of study protocols before funding and start of the study,
 - Regulators and payers need to do away with the firewalls of communication that have existed between them in some healthcare environments, giving way to ongoing dialogue between these communities. This dialogue should ideally result in joint guidance and scientific advice for industry on clinical drug development. While some manufacturers may have mixed feelings about such interactions, it would allow them to factor in, at an early stage of development, the information needs of both regulators and payers, avoiding the need for dual-track development programs. Alignment of the regulators’ and payers’ information needs will have to extend to the pre and post-marketing phases.
 - The increased level of communication between regulators and payers needs to be mirrored in the structure of large companies where (still) existing separations between regulatory affairs departments and groups tasked to address payers’ needs (e.g. managed care marketing groups) may not be sustainable.
 - Regulators will have to more clearly define criteria when RE studies are required for licensing. This may include, for example, life-threatening conditions or situations where placebo-controlled trials or other information leads to uncertainty regarding whether the new drug may be inferior to existing treatment options.

- Both regulators and payers will need to become fully conversant with the complete methodology tool kit for RE assessment. FDA (through the Sentinel initiative) and EMA (through the ENCEPP and IMI initiatives) have already initiated steps to investigate the reliability of observational data in regulatory decision making. As a next step, regulators and payers will need to find a common understanding on the validity and applicability of all available methodologies. It will not be reassuring to patients if payers base coverage decisions on methodology that is judged unreliable by regulators in that same jurisdiction.
- If industry is to be motivated to conduct more post-licensing research and to demonstrate relative (comparative) effectiveness, payers will, in turn, need to be more open to granting some form of conditional reimbursement early after marketing authorisation, with subsequent adjustment of price or reimbursement status, e.g. along the lines of 'coverage with evidence development' developed by the US CMS.