Update on 2004 Background Paper, BP 4 Methods

Table of Contents

1  Overview.............................................................................................................................................. 4

2  Sources of Demographic Data........................................................................................................... 7

3  Sources of Burden of Disease and Mortality Data ........................................................................... 7

  3.1  Burden of Disease and Mortality Analysis...................................................................................... 7

    3.1.1  The WHO Global Burden of Disease 2004 Update ................................................................. 7

    3.1.2  The Lancet Global Burden of Disease Study 2010 ................................................................. 8

4  Country and Regional Definitions Used ......................................................................................... 9

  4.1  The European Union: EU 27 ......................................................................................................... 9

  4.2  The WHO European Region ......................................................................................................... 9

  4.3  The Global Burden of Disease 2010: European regions ........................................................... 9

5  How is the Burden of Disease Measured? ....................................................................................... 10

6  How is Mortality Measured? ............................................................................................................. 10

  6.1  The WHO GBD 2004 Update: Using projections for 2008............................................................. 10

  6.2  The Lancet Global Burden of Disease Study 2010 ...................................................................... 11

7  Risk Factors and How They are Calculated ..................................................................................... 12

  7.1  The WHO 2004 Global Burden of Disease Update ....................................................................... 12

  7.2  The Lancet Global Burden of Disease 2010 Study on Risk Factor Analysis ............................... 12

8  Generating the Burden of Disease and Mortality List .................................................................... 13

  8.1  Using DALY Burden in the Ranking Exercise ............................................................................... 13

  8.2  Using Mortality Burden in the Ranking Exercise ......................................................................... 13

9  Applying Exclusion Criteria to the Burden of Disease and Mortality List ..................................... 13

10 The Preliminary List and Additions to the List .............................................................................. 14

  10.1  Conditions Added Based on Social Solidarity ............................................................................ 14

  10.2  Conditions Added Based on Future Trends and Projections ....................................................... 14

  10.3  Conditions Added Based on Key Risk Factors .......................................................................... 14

11 The Background Reviews ................................................................................................................. 15

  11.1  Questions for the Background Review ........................................................................................ 15

12 Defining a Pharmaceutical ‘Gap’ for Chapter 6 Background Reviews ............................................. 16

13 Use of Clinical Efficacy Reviews for the Analyses of Chapter 6 ................................................... 17

  13.1  Cochrane Reviews ...................................................................................................................... 17

  13.2  Use of “BMJ Clinical Evidence” ................................................................................................. 19
Update on 2004 Background Paper, BP 4 Methods

14 Cochrane Search Strategy

15 Cochrane Reviews: Creating Excel spreadsheets

16 Outcome measure: CDSR Forest plots

17 Outcome Measures: Risk ratios, Odds ratios, and Mean Difference

17.1 Calculating Clinical Efficacy with Relative Risk Ratios

17.1.1 Limitations of Relative Risk Calculations

17.2 Calculating Clinical Efficacy with Odds Ratios

17.3 Calculating Clinical Efficacy with Standardized Mean Difference

18 Modifying the OR and RR Data for Graphical Presentation

19 Limitations of Cochrane Reviews

20 Data sources for R&D funding for “neglected tropical diseases, diarrhoea, pneumonia, tuberculosis, HIV/AIDS and malaria”: the G-FINDER

21 Conclusions

References
1 Overview

The methodology described in this background paper aims to determine pharmaceutical gaps and to create a public health-based research agenda for the EU. In May 2012, a European Commission Working Group meeting was held in Brussels with specific focus to update the Priority Medicines for Europe and the World report (see updated Background paper Chapter 1, Appendix 1.1c). The meeting was hosted by DG Enterprise and Industry. Representatives from Belgium, Hungary, Italy, the Netherlands, Norway, Portugal, Switzerland, and various organizations European Federation of Pharmaceutical Industries and Associations (EFPIA), European Patients Forum (EPF), Entrepreneurship and Innovation Programme (EIP), European Association for Bio-industries (EuropaBio), and the World Health Organization (WHO) were present.

The group agreed that the 2004 Priority Medicines for Europe and the World report “… continues to be relevant in its structure and methodology”. The Working Group decided that the update should follow a dual approach. First, the WHO was tasked with checking whether the therapeutic areas identified in the 2004 report were still appropriate. Second, the WHO was asked to update the 2004 data in Chapters 1 to 6 covering the various priority therapeutic areas. The issues of diagnostics were to be added in those chapters whenever relevant.

Chapter 7 on cross-cutting themes and Chapter 8 on new approaches to promoting innovation were to be updated with a different methodology (see updated Background Chapters 7 and 8).

Reviewers first performed a review of demographic factors (life expectancy, age distribution and the like) for countries in Europe, including EU27 and the world to set the context for the report (see section 2 for data sources and updated Chapter 5 for results). Reviewers then followed the methodology outlined in Figure 4.1.

This Project has combined a number of methods to produce a methodology that can be used for priority setting at country, regional and global levels. The method is intended to be explicit and reproducible (source data are provided on the CD-ROM and website). This chapter provides details of the four complementary approaches used: the evidence-based approach; future projections approach; the risk factor approach; and the social solidarity approach.
In broad outline the methodology is as follows:

a. The first step was a review of demographic factors (such as life expectancy and age distribution) for countries in Europe (including the EU27) and the world to set the context for the Report.

b. A ranking exercise was then carried out, using burden of disease information (disability adjusted life years (DALYs) and mortality), to generate two lists, one list of the major diseases and conditions which account for the majority of the total DALY
burden in both the EU 27 and the rest of the world and the second, counterpart list for the total mortality burden of major diseases and conditions (see Tables 5.3.1 and 5.4.1 in Chapter 5). These are called the burden of disease and mortality lists in Figure 4.1.

c. Some of the conditions on these lists, such as road traffic accidents, were then excluded, mainly because pharmacotherapies were not amenable to deal with these conditions.

d. Additional criteria derived from the three other approaches (Section 7) were then applied to generate additions to the conditions on these two lists. These included: health-related projections and trends; risk factors; and social solidarity/social justice/equity (see Section 6). Combination of the DALY and mortality lists, removal of duplicate conditions, and additions based on the three other approaches, were combined to generate the Preliminary List (Figure 4.1).

e. Each condition and disease on this Preliminary List was reviewed in an in-depth Background paper (see Section 11). These reviews can be found in Chapter 6 and associated Annexes and Appendices. The primary purpose of these Background reviews was to come to a conclusion as to whether or not a pharmaceutical gap existed for that particular condition. Many of these reviews looked at evidence for clinical efficacy using the Cochrane Collaboration (see Section 13.1) and other databases such as BMJ Clinical Evidence (see Section 13.2). The final list of priority disease/conditions and their associated pharmaceutical gaps is found in Chapter 9 of this report.

The remainder of this Background document provides more detail on the methodology. Sections 2 and 3 discuss the various sources of demographic and burden of disease and mortality data, respectively. Section 4 summarizes the countries that consist of various European regions; among the different data sources, there is no consistent definition of a European region. Sections 5 and 6 briefly summarize how burden of disease and mortality were measured in the various databases. Section 7 discusses how risk factors were measured in the databases, and Section 8 summarizes how the burden of disease and mortality list is generated based on a ranking system (see Figure 4.1). Section 9 discusses how this burden of disease and mortality list was further screened using exclusion criteria. These results and other considerations generated the Preliminary List for Section 10.

Once the Preliminary List was created, each condition on this list became the subject of a background review (see completed reviews in Chapter 6 and associated documents).

Sections 11-12 discuss what is intended to be included in these Background papers, including the operational definition of a pharmaceutical gap.

Sections 13-19 describe in detail the strategy and metrics used to in some of the Background reviews to determine clinical efficacy of medicines to treat the conditions on the Preliminary List. The methods are based on analyses of Cochrane Reviews of clinical trials. These sections also describe a method devised to graphically project some of this data, as well as limitations of the method and Cochrane database. Section 20 is the conclusion.
2 Sources of Demographic Data

The demographic component of this report is based on important regional and international reports and databases in which reviewers analysed the following parameters: life expectancy at birth, age distribution of the world population, fertility rates and distribution of people living in urban and rural areas. The primary database used was the World Development Indicators database. See Background Document Chapter 5.

3 Sources of Burden of Disease and Mortality Data

Information needed to derive the burden of disease and mortality List (see Figure 4.1) was based on public databases developed by the WHO Global Burden of Disease 2004 Update (see Section 3.1.1), as well as on a recent series of documents provided by the Lancet Global Burden of Disease Study 2010 (see Section 3.1.2).

3.1 Burden of Disease and Mortality Analysis

A ranking exercise was conducted using information based on DALY and mortality for countries of Europe and the world. In brief, a list was generated of the conditions that account for the top 20 of the total global burden of disease. For the WHO European Region, reviewers took the top 20 conditions based on DALY burden of disease. For both the world and EU 27 countries, we generated a list of the top 10 conditions based on mortality burden. Several databases were used to develop this ranking exercise.

3.1.1 The WHO Global Burden of Disease 2004 Update

The WHO Global Burden of Disease (GBD) project is a systematic, scientific effort to quantify the comparative magnitude of health loss due to diseases, injuries, and risk factors by age, sex, and geographies for specific points in time. The GBD produces results on mortality, cause-specific mortality, years of life lost (YLLs), years lived with disability (YLDs), and disability-adjusted life years (DALYs). In 1990, the GBD was assessed by the WHO for eight regions of the world. The methodology was partially revised and further analyses were done by the WHO for 1999-2002 and 2004 for 14 regions. It provided detailed global and regional estimates of premature mortality, disability, and loss of health for 135 causes by age and sex, drawing on extensive WHO databases and on information provided by Member States.

The WHO GBD 2004 update provided estimates for mortality, incidence, prevalence, YLLs, YLDs, and DALYs by age, sex, and cause. It also provided projection of these estimates for the years 2008, 2015, and 2030. For the purposes of this updated Report, reviewers used the WHO projections for DALYs and mortality for 2008. This database makes use of the ICD (International Classification of Diseases) to group diseases into bigger clusters, such as ‘Ischaemic heart disease’. To generate the “burden of disease and mortality” list, reviewers made use of the most specific categories given by this WHO database.
3.1.2 The Lancet Global Burden of Disease Study 2010

The Bill & Melinda Gates Foundation provided funding for a new round of the Global Burden of Disease study towards a multi-centre collaboration to revise and update global estimates of burden of disease, injury, and risk factors for the years 1990, 2005, and 2010. The study was led by the Institute for Health Metrics and Evaluation (IHME) at the University of Washington and other institutions including the WHO, Harvard University, Johns Hopkins University, and the University of Queensland. High level summary papers with final estimates and methods were published in the Lancet journal as of 14 December 2012. Up until the publication of this 2010 study, there had not been a comprehensive and systematic revision of the GBD since the WHO GBD project in 2004 (see Section 3.1.1).

The Lancet GBD 2010 study provided estimates of disease burden and mortality for 20 age-groups, both sexes, and 21 regions for years 1990 and 2010. Analyses from the Lancet GBD 2010 study differ from the WHO GBD 2004 study in several ways. The Lancet study used a new standard life table with a life expectancy at birth of 86.0 years. Years lived with disability were computed using prevalence and disability weights and took comorbidity into account. Years of life lost, YLDs, and DALYs were computed with no discounting of future health and no age-weights. Death rates and numbers have been estimated with 95% uncertainty intervals (UIs). This implies a substantial shift towards greater weight being given to deaths at younger ages, especially younger than 5 years, and greater weight to deaths compared to non-fatal health loss. A detailed summary of their general methodology is beyond the scope of this report, but various papers produced and published in the literature can be found there.

In some areas, the Lancet GBD 2010 study results are broadly similar to the WHO’s recent estimates. In other ways, the results of the Lancet GBD 2010 study differ substantially from existing analyses done by the WHO and other United Nations agencies at global, regional, and country levels. The methodology of the Lancet GBD 2010 study is different in some respects from the earlier WHO methodology. The methodology in the 2004 WHO Priority Medicines Report was based on the earlier WHO methodologies and on changes in global health. For example, the geographic definitions differ between the WHO GBD 2004 update and the Lancet GBD 2010 study (see below Section 4 et seq.). The Lancet GBD 2010 study introduced new methods for analyzing mortality data on child and adult survival, several new model life tables, new methods for data synthesis, the Cause of Death Ensemble model (CODEm), new methods for collecting and analyzing data on assessments of disability, and new methods for estimating risk factor trends. Both the WHO GBD 2004 and the GBD 2010 methods are open to challenge. The WHO intends to convene an expert consultation in February 2013 to review all current work on global health estimates and discuss ways to standardize and improve data and estimates for better country, regional, and global health decision making.5
4 Country and Regional Definitions Used

4.1 The European Union: EU 27

In the original 2004 Priority Medicines Report, and in some parts of Chapter 6 of this report, reviewers use data on the EU 25-27 countries listed below, as well as on subunits of the EU27, depending on when various countries joined the EU:

EU 27: (The 27 Member States of the European Union in 2013) Austria, Belgium, Bulgaria, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and the United Kingdom.

Where data on the 27 individual EU countries was not available at all, reviewers used data on the WHO European Region (see Section 4.2). This was the case, for example, for DALY statistics that were used based on the WHO GBD 2004 update provided by the WHO.

4.2 The WHO European Region

The WHO European Region comprises over 50 countries, covering a vast geographic region from Iceland to Kazakhstan. The WHO European Region is much larger than the 27 countries of the EU:

Andorra, Armenia, Austria, Azerbaijan, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Luxembourg, Malta, Monaco, Netherlands, Norway, Poland, Portugal, Republic of Moldova, Romania, Russian Federation, San Marino, Serbia and Montenegro, Slovakia, Slovenia, Spain, Sweden, Switzerland, Tajikistan, the former Yugoslav Republic of Macedonia, Turkey, Turkmenistan, Ukraine, United Kingdom, Uzbekistan.

4.3 The Global Burden of Disease 2010: European regions

The Lancet GBD 2010 study used different regions that were based on criteria of epidemiological homogeneity and geographic contiguity. A total of 21 regions were created of which three are relevant to the European countries:

1) Central Europe:
   Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Hungary, Macedonia, the Former Yugoslav Republic of Montenegro, Poland, Romania, Serbia, Slovakia, Slovenia.

2) Eastern Europe:
   Belarus, Estonia, Latvia, Lithuania, Moldova, Russian Federation, Ukraine.

3) Western Europe:
   Andorra, Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, Malta, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom.

If reviewers used any or all of these regions when discussing results on burden of disease and mortality (see Chapter 5) it will be specified which geographic definitions were used.
5 How is the Burden of Disease Measured?

The WHO has defined ‘burden of disease’ as the significance of a particular disease for society beyond the immediate cost of treatment. Burden of disease measurements are based on a comprehensive and consistent set of estimates of mortality and morbidity by age, sex, and geographic region. The primary quantifier of burden of disease is the disability adjusted life year (DALY).

The DALY is a summary measure of population health that combines in a single indicator years of life lost from premature death and years of life lived with disabilities. One DALY can be thought of as one lost year of ‘healthy’ life, which contributes to the burden of disease as a measurement of the gap between current health status and an ideal situation where everyone lives into old age free of disease and disability. DALYs are a way of aggregating the number of life years lost by people suffering from each disease with the amount of disability suffered while they are still alive. These two amounts are combined in a complex manner to give the overall burden of that disease. Disease burdens are thus measured in DALYs lost due to each disease.

Consistent and meaningful DALY estimates depend on a clear definition of the condition under consideration in terms of case or episode and severity level or disease stage. It is then necessary to ensure that the disability weight and the population incidence or prevalence data relate to the same case definition. Thus far, there appears to be a scarcity of empirical data on health state valuations. The DALY concept has generated a debate and a literature of its own within the public health community.

The burden of disease approach can be broken down to show the relative contributions of different conditions to the overall burden of disease, and it can show the burden of disease that can be attributed to known risk factors. A risk factor analysis was used in the methods as well (see Section 10.1).

The global burden of disease methodology of this report, although not perfect, is the best single tool available for the intended audience, of strategic planners and decision makers. It provides a single summary measure of ill health and a fundamental tool for policy makers when considering the relative benefits of different policy options.

6 How is Mortality Measured?


The WHO GBD 2004 update contains projected summary estimates of mortality for the WHO Member States for the year 2008, although the WHO finished the analysis in 2010. Data, methods, and cause categories are described in an accompanying text document available on the WHO website.

The estimates were classified by regions, age groups, and conditions based on the WHO burden of disease classification. Apart from the inclusion of new epidemiological data for
specific causes, the 2008 mortality estimates have incorporated more recent vital registration (VR) data for many countries and VR data for a number of countries for the first time. Furthermore, the database now has updated additional information on levels of child and adult mortality and improvements in methods used for the estimation of causes of child deaths in many countries without good death registration data.

The WHO contacts member states directly on a routine basis to obtain their latest data on cause of death from their vital registration sources. If countries do not have a complete or accurate registration system, they are requested to send in data from other reliable sources, such as population surveys, epidemiological studies, and health facility data. Mortality estimates were thus based on analysis of latest available national information on levels of mortality and cause distributions together with the latest existing information from WHO programs, the International Agency for Research on Cancer (IARC), and United Nations programme on HIV/AIDS (UNAIDS) for specific causes of public health importance. All received data on cause of death are then divided into three main groups of causes:

1) Communicable diseases, maternal causes, conditions arising in the perinatal period, and nutritional deficiencies.
2) Non-communicable diseases.
3) Intentional and unintentional injuries.

Within these main groups, all diseases from the International Classification of Diseases (ICD) are represented. A more detailed discussion of the methods of obtaining mortality data for the update of the WHO GBD 2004 analysis can be found in http://www.who.int/healthinfo/global_burden_disease/cod_2008_sources_methods.pdf

6.2 The Lancet Global Burden of Disease Study 2010

The Lancet Global Burden of Diseases, Injuries, and Risk Factors Study 2010, aimed to estimate annual deaths for the world and 21 regions between 1980 and 2010 for 235 causes, with uncertainty intervals (UIs), separately by age and sex. All available data on causes of death for 187 countries for 2010 was extracted from vital registration, verbal autopsy, mortality surveillance, censuses, surveys, hospitals, police records, and mortuaries.

Data quality was checked for completeness, diagnostic accuracy, missing data, stochastic variations, and probable causes of death. Six different modeling strategies were used to estimate cause-specific mortality trends depending on the strength of the data. For example, for 133 causes of death, the Cause of Death Ensemble model (CODEm) approach uses four families of statistical models testing a large set of different models using various permutations of covariates (model ensembles were developed from these component models). For selected causes (African trypanosomiasis, congenital syphilis, whooping cough, measles, typhoid and parathyroid, leishmaniasis, acute hepatitis E, and HIV/AIDS) models based on incidence, prevalence, and case-fatality were used. A more detailed discussion of the methods of obtaining global and regional mortality data is beyond the scope of this report but can be found in the Lancet GBD Study 2010 by Lozano et al. A more detailed discussion of the methods of obtaining age and gender-specific mortality data can be found in the Lancet GBD Study 2010 by Wang et al.
7 Risk Factors and How They are Calculated

7.1 The WHO 2004 Global Burden of Disease Update

In addition to burden of disease and mortality ranking, this update includes risk factors (e.g. smoking and obesity) as a measure of disease burden.

In brief, the current burden of disease and injury attributable to health risk is estimated by calculating the proportional reduction in population disease or mortality that would occur if exposure to a risk factor were reduced to some alternative ideal exposure scenario (i.e. the counterfactual). To improve comparability across risk factors, a counterfactual distribution was defined for each risk factor as the population distribution of exposure that would lead to the lowest levels of disease burden. For some risks such as smoking, the level of exposure is clear - zero tobacco use is the ideal. For others, like blood pressure, the distribution of exposure leading to the lowest levels of disease burden is not obvious. Many people have a clinically normal range for blood pressure (i.e. below 140 mmHg), but have blood pressure levels above ideal levels. For blood pressure, the WHO selects a blood pressure at the low end of the normal range (see WHO Surveillance of Risk Factors Reports). 21,22

Of course, many diseases are caused by multiple risk factors that may interact with one another. For example, two risk factors – smoking and urban air pollution – cause lung cancer. Some lung cancer deaths are attributed to both smoking and air pollution. For lung cancer, deaths that could have been averted by removing or lowering either tobacco exposure or urban air pollution is the area of overlap of the respective counterfactual distributions. Risk factors identified by this WHO GBD 2004 Update were used to help generate the Preliminary List (see Section 10).

7.2 The Lancet Global Burden of Disease 2010 Study on Risk Factor Analysis

A key aspect of the Lancet Global Burden of Disease (GBD) 2010 study was to calculate the proportion of deaths or disease burden caused by specific risk factors (i.e. ischaemic heart disease caused by increased blood pressure) holding other independent factors unchanged. These calculations were done for 20 age groups, both sexes, and 187 countries for 2010. 17 Distributions of exposure to risk for each year, region, sex, age group, and relative risks per unit of exposure were estimated by systematically reviewing and synthesizing published and unpublished data. These estimates, together with estimates of cause-specific deaths and DALYs from the Lancet Global Burden of Disease Study 2010, were used to calculate the death and DALY burden attributable to each risk factor exposure compared with some theoretical-minimum-risk exposure. A more detailed discussion of the methods of determining deaths from DALYs and sum of YLDs and YLLs attributable to the independent effects of 67 risk factors and clusters of risk factors can be found here. 23
8 Generating the Burden of Disease and Mortality List

In the ranking of high burden diseases from which background reviews were generated, reviewers made use of both burden of disease DALYs and mortality data. The primary focus was on burden of disease as disability, recognizing the burden that disease can put on someone’s life without causing direct death. Pharmaceuticals that can help people live in a fulfilling way are viewed as being just as important as lifesaving pharmaceuticals. Using both the WHO GBD 2004 update and the Lancet GBD 2010 study, reviewers recorded the top 20 diseases with the highest DALY burden for Europe (WHO European Region) and the world (all countries including WHO European Region). Moreover, the top 10 diseases with the highest mortality for both Europe and the world was recorded.

8.1 Using DALY Burden in the Ranking Exercise

Since the WHO GBD 2004 update and the Lancet GBD 2010 analyses contain information on different regions in the world, reviewers could not directly compare outcomes for specific countries of the European Union with outcomes of the world as a whole. Diseases were ranked by DALY burden based on those specific causes of disease burden found in WHO GBD 2004 update (using 2008 estimates). The ranking was done for both the WHO European Region (Section 4.2) and the world (including the WHO European Region) after which both rankings were combined into the burden of disease and mortality list (Figure 4.1).

8.2 Using Mortality Burden in the Ranking Exercise

For rankings of conditions based on mortality, reviewers also used the 2008 mortality estimates from the WHO 2004 GBD update data for Europe (EU27 countries) and the world (including the EU27). Acute conditions with high mortality, but not necessarily high DALY burden, would be found by this ranking exercise, and the top 10 mortality conditions for countries in the EU27 and the world were included into the burden of disease and mortality list.

9 Applying Exclusion Criteria to the Burden of Disease and Mortality List

From the burden of disease and mortality list generated in Section 8, reviewers then eliminated conditions or diseases that, based on reviewers’ experience and literature review, has no known pharmaceutical interventions designed to cure or treat the specific condition. The eliminated conditions include: intentional and unintentional injuries, road traffic accidents, refractive errors, birth trauma, and childhood cluster diseases.

More importantly, the disease groups in both the WHO GBD 2004 and the Lancet GBD 2010 analyses contained grouped categories of many other less-impact specific diseases (such as ‘other malignant neoplasms’, ‘other digestive diseases’ and so on), which were also excluded from the list. The exclusion of these other categories of diseases was done for two reasons. First, it would not have been feasible to search for all the individual diseases contributing to
the cluster within the time frame of this project. Second, it should be noted that the individual burdens of the specific diseases contributing to the cluster are all relatively small, but the very fact that there are many individual diseases in the cluster is what makes disease cluster end up high in the ranking.

10 The Preliminary List and Additions to the List

The net result of this screening (see Section 9) contributes to the Preliminary List of diseases and conditions in Chapter 5. Three other domains were analysed in order to allow us to add other conditions to the Preliminary List. One domain was risk factors based on existing analyses done by the WHO GBD 2004 work (see Section 10.3). Additional diseases were added that were based on reviewers’ judgment and considerations of two other domains: social solidarity and future trends and projections (demographic and epidemiologic).

10.1 Conditions Added Based on Social Solidarity

Reviewers have chosen to emphasize the ethical and moral aspects of priority setting as another prioritization method. Many European countries have a long history of social solidarity. This has been demonstrated by the creation of universal social security systems and national health systems, which ensure that everyone in that society should have access to medical care and pharmaceuticals. The USA and Europe have chosen to pass legislation for orphan diseases, which requires the society to spend substantial funds on a limited number of afflicted individuals who suffer from rare diseases. At a global level it is suggested that, based on principles of global solidarity, similar efforts should be made to cure neglected diseases that afflicted poor people in underdeveloped countries, as well as conditions affecting women and children, such as postpartum hemorrhage and maternal mortality.

10.2 Conditions Added Based on Future Trends and Projections

What are the emerging diseases that will affect the European Union and the world going forward? What existing diseases will grow in importance? The answers to these questions form a third prioritization method and are based primarily on consensus judgments and observational and clinical evidence. Resolutions of the WHO World Health Assembly (WHA) and the European Parliament have identified antibacterial resistance (AMR) as a serious threat to global public health. \(^{24, 25}\) Antibacterial resistance (see Background paper 6.1) and pandemic influenza (see Background paper 6.2) are an important threat to global public health that will continue to grow, which are sufficient reasons to include these topics in this report.

10.3 Conditions Added Based on Key Risk Factors

What are the growing risk factors that will affect the European Union and the world going forward? The answers to these questions are based on data generated by the WHO GBD 2004 update and the Lancet GBD 2010 analyses. Cessation of smoking and combating obesity are
two risk factors amenable to pharmacotherapeutics and are added to the Preliminary List for that reason. Alcohol consumption is another risk factor addressed in this report.

11 The Background Reviews

The primary purpose of these background reviews is to determine if the condition on the Preliminary List contains a pharmaceutical gap, as defined below in Section 12. For all conditions on the Preliminary List, questions were asked modified from the template developed by the Global Forum on Health Research. Each author of the background review was free to answer these questions in any order they wish.

The two databases (the WHO GBD 2004 Update and the Lancet GBD 2010) both used the International Classification of Diseases (ICD) to aggregate different very specific diseases into more general terms (such as ‘cerebrovascular disease,’ ‘ischemic heart disease,’ or ‘osteoarthritis’). Since these are clustered terms, the ICD classified diseases that correspond to those in the Preliminary List for the purposes of guiding analysis in the respective background reviews were used.

11.1 Questions for the Background Review

What is the size and nature of the disease burden?
- What are the epidemiological trends?
- What are the current or likely future factors that impact on disease burden at the following levels, and in what way (e.g. individual, community and household)?
  - Health sector (health ministry, systems and service delivery)
  - Non-health sectors
  - Government and international

What is the control strategy? Is there a pharmaceutical gap?
- Is there an effective package of control methods assembled into a control strategy for most epidemiological settings including therapeutic prevention?
- What are its current components (stratify by geographical areas if necessary)?
- What is known of the affordability, feasibility, and sustainability of the control strategy?
- What is known about the clinical efficacy of the current components? (See below)
- If such a control strategy exists, how effective is it based on observation, or could it be based on epidemiological modeling at:
  - reducing morbidity
  - preventing mortality
  - reducing transmission
- Why does the disease burden persist?

What can be learnt from past/current research into pharmaceutical interventions for this condition?
- From past or current research?
- What is known about existing research resource flows?
What is the current ‘pipeline’ of products that are to be used for this particular condition?
- What is known, if anything, on the safety and efficacy of products in the pipeline?

What are the opportunities for research into new pharmaceutical interventions?
- What is the state-of-the-art science (basic and operational) for this disease and what opportunities does it offer?
- What is the current status of institutions and human resources available to address the disease?

What are the gaps between current research and potential research issues which could make a difference are affordable, and could be carried out in a) five years or b) in the longer term? For which of these gaps are there opportunities for pharmaceutical research?
- Which issues can only be realistically addressed with increased financial support or investment in human and institutional capacity?
- Which issues are best suited to the comparative advantage of the EU?

12 Defining a Pharmaceutical ‘Gap’ for Chapter 6 Background Reviews

Figure 1.1 in Chapter 1 illustrates the conceptual framework for looking at pharmaceutical ‘gaps’. The concept will be more comprehensively defined for purposes of in-depth Background reviews of Chapter 6. The nature of a pharmaceutical ‘gap’ depends both on the kind of disease and on several critical domains that are particularly important for these diseases. The following Table 4.1 lists the main category of disease in this analysis of ‘gaps’ and the three ‘domains’- written as a series of questions. In each box, conditions which need to exist in order for there to be a pharmaceutical gap are defined.

For the purpose of the Background reviews, each author was asked to determine whether or not available evidence suggested that treatments were efficacious. Analyses of clinical efficacy as described below were used.
Table 4.1: Operational definition of pharmaceutical ‘gaps’ for purposes of this report

<table>
<thead>
<tr>
<th>DISEASE</th>
<th>DEVELOPMENT OF RESISTANCE?</th>
<th>IMPORTANCE OF INDIVIDUAL VARIATION IN CLINICAL RESPONSE?</th>
<th>IS THE TREATMENT A COMPLETE CURE? i.e., ARE THERE ONLY PALLIATIVE MEDICINES?</th>
<th>DO WE UNDERSTAND THE BASIC MECHANISMS OF DISEASE?</th>
</tr>
</thead>
</table>
| Communicable, infectious     | 1. A gap exists if there are less than 4 efficacious therapies.  
2. A gap exists if the infection-causing agent has a complex life cycle and the intervention(s) are directed at only ONE part of the life cycle | N/A                                                      | A gap exists if there is no complete cure                                    | N/A                                               |
| Chronic, non-communicable    | N/A                        | A gap exists if there are less than 4 efficacious interventions, as this does not take individual variation into account** | A gap exists if there is no complete cure e.g., COPD                       | A gap exists if we do not yet know the how various factors (biological/environmental) actually cause the particular condition to present in humans |
| Genetic deficiency           | N/A                        | N/A                                                      | A gap exists if no single intervention is sufficient to correct a metabolic/genetic deficiency that causes a disease | N/A                                               |

13 Use of Clinical Efficacy Reviews for the Analyses of Chapter 6

13.1 Cochrane Reviews

Because the key output of this project is a determination of pharmaceutical gaps, it is useful to have a measure of the clinical efficacy or lack of efficacy for pharmaceutical interventions. For most diseases and clusters identified from the Preliminary List, reviewers looked at the Cochrane Database of Systematic Reviews for information regarding the clinical efficacy of any new pharmacological interventions since the original 2004 Report to treat these conditions. Reviewers also looked at the British Medical Journal (BMJ) Clinical Evidence reviews as a supplemental source of information (see Section 13.2).
Update on 2004 Background Paper, BP 4 Methods

The Cochrane Database of Systematic Reviews (CDSR) are highly structured literature reviews that synthesize results from the highest quality research studies, usually randomized, placebo-controlled trials. The data is compiled and evaluated using explicit criteria, and the results are combined statistically. Extensive literature searches are conducted to identify relevant studies from the international medical literature. The work of the Cochrane groups is considered the gold standard in the search for the best systematic reviews of medical evidence.

Clinical efficacy is a measure of the accuracy or success of a diagnostic or therapeutic technique when carried out in a clinical trial. The Cochrane Reviews analysed should be viewed in terms of clinical efficacy and not clinical effectiveness; the latter defined as the accuracy or success of a diagnostic or therapeutic technique when carried out in an average real-world clinical environment. That is, clinical effectiveness is the extent to which a treatment achieves its intended purpose. Reviewers will not use these terms interchangeably and the reader should be aware of this distinction. If the distinction is not important, it will be noted this.

Highly structured analyses of the efficacy of clinical pharmaceutical interventions can be traced back to at least the early part of the last century when Karl Pearson began to use formal techniques to combine data from different health studies. Later, in 1979, the British physician and epidemiologist, Archie Cochrane, introduced the idea of creating reliable reviews of the available clinical evidence based upon randomized, controlled clinical trials.

The Cochrane Database of Systematic Reviews (CDSR) comes from the international Cochrane Collaboration. The international Cochrane Collaboration is an organization of over 7,000 health professionals, researchers, scientists and consumers from approximately 80 countries. There are 50 international Cochrane review groups. Cochrane reviews have already provided a substantial amount of evidence relevant to health care. They are relevant to a wide range of policies and are widely cited in guidelines. Their importance was recognized in an early report from the American Institute of Medicine. Journals such as the British Medical Journal (BMJ) now recommend that their editorialists should refer to relevant Cochrane reviews where these exist.

Although it is useful to systematically review a body of data, it may be misleading to statistically pool results from separate studies. For instance, in case-control studies confounding and bias may produce estimates of associations that deviate from the true causal effects beyond what can be attributed to chance. Combining a set of epidemiological studies in this situation will thus often provide spuriously precise, but biased, estimates of associations. However, randomized, controlled trials, which are the primary source of information for CDSR, generally provide an unbiased estimate of the underlying treatment effect. The overall effect of various individual analyses is obtained by combining the data. Reviews in the CDSR use a weighted average of the results where the larger trials have more influence than the smaller ones; results from small studies are more subject to effects of chance and should therefore be given less weight. Cochrane Reviews do not exist for all conditions, but there has been a major increase in the number of Cochrane Reviews since 2004 as shown in Figure 4.2.
Another drawback is that most trials data in the Cochrane Database are randomized placebo-controlled trials and do not include other data such as that from observational studies. (See also Chapter 8.4).

The difficulty faced in obtaining this data underlines the need to improve the regulatory process using Phase IV research (post-marketing surveillance) to collect data on both adverse events and exposure to pharmaceuticals in large numbers of patients.

Figure 4.2: Cochrane Database of Systematic Reviews: Total Cochrane Reviews and Protocols

Source: Cochrane Database at http://www.cochrane.org/cochrane-reviews/cochrane-database-systematic-reviews-numbers

13.2 Use of “BMJ Clinical Evidence”

There are a series of systematic reviews in tabular form that summarize the current state of knowledge and uncertainty about the prevention and treatment of clinical conditions based on thorough searches and appraisal of the literature. In effect, they summarize systematic reviews (including Cochrane Reviews), RCTs, and observational studies where appropriate. Clinical evidence is directed to clinicians as a way of allowing them to know when their uncertainty stems from gaps in the evidence rather than gaps in their own knowledge. We have, therefore, supplemented evidence gleaned from Cochrane Reviews with summary reports developed from BMJ Clinical Evidence.23
14 Cochrane Search Strategy

Using the Preliminary List, the Cochrane library (www.thecochranelibrary.com) was searched for the diseases listed as causes or ICD names. Using ICD names, CDSR was searched for “ICD name 1” OR “ICD name 2” and so on using just the “TITLE” and “KEYWORDS”. This CDSR search was limited to only those reviews found using this search strategy. Specifically, PubMed MeSH terms were used to find synonyms to the name of the disease and then a search performed with those words in title or in keywords. For example (depression* OR “depressive episode*” OR “depressive disorder”) in title or keywords. All searches were limited to only current Cochrane Reviews and not to withdrawn reviews or to mere protocols. After searching Cochrane, BMJ Clinical Evidence was reviewed for more information on treatments that were not mentioned in Cochrane.

All CDSR studies published after 2004 that had pharmaceutical interventions using placebo-controlled, randomized trials were included. In addition, relevant reviews dealing with head-to-head comparative studies (e.g. between two medicines) were considered on a disease basis.

For a given disease condition, there are often several Cochrane reviews, which represent different interventions, outcomes, dosage forms, and numbers of trial participants. Further, not all conditions or ICD codes were searched in the Cochrane Database and this was left up to the individual reviewer.

15 Cochrane Reviews: Creating Excel spreadsheets

For each Cochrane Review for a given condition or ICD disease, a standard data collection sheet was constructed using Excel. Data was collected on:

- Type of review: Cochrane or BMJ
- Title
- Reference (can be found under ‘how to cite’ in Cochrane)
- Patient group (what were the characteristics of the participants)
- Intervention used
- Control group used
- Outcome event (this is always more than one such outcome of the clinical trials e.g. death, number of visits to emergency room, et al.)
- Outcome measure for that particular outcome event which was relative risk or odds ratio or mean difference (often used for neuropsychiatric conditions like depression)
- Number of participants and number of trials for that particular outcome measurement
- Outcome measure (the value of relative risk, odds ratio)
- Conclusion (authors conclusions, briefly in words)
- Lower confidence interval value of outcome measure
- Upper confidence interval value of outcome measure
- Adverse events reported
Outcome measure: CDSR Forest plots

Forest plots—the graphical display of individual study results and, usually, the weighted average of studies included in a systematic review—are one way of summarizing the review’s results for a specific outcome (Figure 4.3). Plots of this kind first appeared in the 1970s and were refined over the next two decades; they were first called ‘forest plots’ in the mid-1990s. Since that time the elements contained in a forest plot and the layout of such plots have become somewhat standardized, largely due to the introduction of software that helps authors construct these plots.

One common criticism of this type of meta-analysis is that it sums up the results from different studies and calculates a summary statistic as if it is one big study. This concept is faulty and a meta-analysis does not follow this method. Instead, a meta-analysis looks at the results within each study and calculates a weighted average, which reflects the actual numbers in each study.

Figure 4.3: Forest plot of comparison: 1 All placebo-controlled studies, Outcome: At least 50% pain reduction over baseline.

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Gabapentin Events</th>
<th>Total</th>
<th>Placebo Events</th>
<th>Total</th>
<th>Risk Ratio M.H. Fixed, 95% CI</th>
<th>Risk Ratio M.H. Fixed, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1 Postherpetic neuralgia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irving 2009</td>
<td>29</td>
<td>107</td>
<td>8</td>
<td>51</td>
<td>4.0%</td>
<td>2.30 [1.63, 3.20]</td>
</tr>
<tr>
<td>Rice 2001</td>
<td>74</td>
<td>223</td>
<td>18</td>
<td>111</td>
<td>10.0%</td>
<td>2.30 [1.41, 3.73]</td>
</tr>
<tr>
<td>Wallace 2010</td>
<td>95</td>
<td>268</td>
<td>36</td>
<td>131</td>
<td>24.0%</td>
<td>1.29 [0.93, 1.77]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>599</td>
<td>939</td>
<td>293</td>
<td>386</td>
<td>16.0%</td>
<td>1.07 [1.29, 2.16]</td>
</tr>
<tr>
<td>Total events</td>
<td>188</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneity Chi² = 4.80, df = 2 (P = 0.09), I² = 50%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect Z = 3.90 (P &lt; 0.0001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2 Painful diabetic neuropathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backonja 1998</td>
<td>39</td>
<td>84</td>
<td>10</td>
<td>81</td>
<td>8.1%</td>
<td>2.35 [1.49, 3.86]</td>
</tr>
<tr>
<td>CTR 945, 1998</td>
<td>55</td>
<td>156</td>
<td>19</td>
<td>77</td>
<td>12.9%</td>
<td>1.42 [0.81, 2.44]</td>
</tr>
<tr>
<td>CTR 245, 224</td>
<td>77</td>
<td>223</td>
<td>48</td>
<td>179</td>
<td>23.5%</td>
<td>1.39 [0.13, 2.15]</td>
</tr>
<tr>
<td>Peraz 2000</td>
<td>14</td>
<td>47</td>
<td>2</td>
<td>16</td>
<td>11.1%</td>
<td>1.12 [0.87, 2.20]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>147</td>
<td>360</td>
<td>20</td>
<td>136</td>
<td>45.5%</td>
<td>1.78 [1.43, 2.18]</td>
</tr>
<tr>
<td>Total events</td>
<td>166</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneity Chi² = 0.27, df = 3 (P = 0.69), I² = 0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect Z = 0.14 (P = 0.9009)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3 Mixed neuropathic pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sarwell 2002</td>
<td>32</td>
<td>151</td>
<td>22</td>
<td>152</td>
<td>15.9%</td>
<td>1.45 [0.93, 2.27]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>153</td>
<td>452</td>
<td>10.9%</td>
<td>145</td>
<td>1.45 [0.98, 2.37]</td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>32</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneity: Not applicable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect Z = 1.46 (P = 0.14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.4 Nerve injury pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orenth 2008</td>
<td>13</td>
<td>98</td>
<td>9</td>
<td>98</td>
<td>4.5%</td>
<td>1.44 [0.96, 2.22]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>98</td>
<td>98</td>
<td>4.5%</td>
<td>98</td>
<td>1.44 [0.96, 2.22]</td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>13</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneity: Not applicable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect Z = 0.30 (P = 0.77)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1.5 Small fibre sensory neuropathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ho 2005</td>
<td>5</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>0.5%</td>
<td>5.00 [0.95, 30.65]</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>10</td>
<td>10</td>
<td>0.5%</td>
<td>10</td>
<td>5.00 [0.95, 30.65]</td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneity: Not applicable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect Z = 1.94 (P = 0.12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>1335</td>
<td>203</td>
<td>100.0%</td>
<td>1.70 [1.46, 2.19]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>436</td>
<td>173</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneity Chi² = 12.03, df = 3 (P = 0.017), I² = 20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test for overall effect Z = 6.74 (P = 0.0001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Taken from Moore et al. Gabapentin for chronic neuropathic pain and fibromyalgia in adults, published online March 2011
17 Outcome Measures: Risk ratios, Odds ratios, and Mean Difference

The definition of a pharmaceutical ‘gap’ (see Table 4.1) and thus the inclusion criteria for the Cochrane Database requires information on the results of clinical trial studies in which patients are randomised to either treatment or placebo or comparative product, followed for a fixed amount of time, and are observed to experience a response to treatment (event/no event such as death, whether the patient visited the emergency room within one week, recurrence of the condition within a certain time period, and so on). The effect of an intervention is measured by comparing the events in the two groups of patients. The outcomes that were primarily analysed were relative risk ratios (RR), odds ratios (OR), and standardized mean differences (SD).

17.1 Calculating Clinical Efficacy with Relative Risk Ratios

In Figure 4.3, the first column has the study name; the second column the proportion of trial participants receiving gabapentin that experienced at least 50% pain reduction over baseline; and the third column shows the proportion of trial participants receiving placebo that experienced at least 50% pain reduction over baseline. In the Forest plot showing a relative risk diagram, the vertical line indicates a relative risk of 1.0, which indicates a situation where the treatment and placebo effects are equivalent.

The horizontal dots and bars represent the relative risk and a 95% confidence interval for each study. That is, if 100 trials for each study were performed, 95 of them would show treatment effects between the upper and lower boundaries. The column labelled “weight” is proportional to the statistical variability of the study; high variability, associated with a small study population, means less weight is given to that study, and vice versa. The relative risk is shown numerically in the final column of each table.

The weighted summary estimate of relative risk, that is the mean effect of all the studies (“Total”) with its own confidence interval is shown in the last row, along with an estimate of how the individual trials from vary from each other (the chi-square). Thus, if all the trials were repeated 100 times, the weighted estimate would show treatment effects between the upper and lower boundaries. In this example, the summary statistics of the confidence intervals of relative risk (“Total” row) far exceed 1 (1.46 to 1.99), suggesting that the association of treatment with at least a 50% reduction in pain in outcome measure is statistically significant compared to placebo. Thus, there is enough evidence that the association between gabapentin administration and pain reduction is significant.

If the lower boundary of the confidence interval had included a value of 1, reviewers would only be able to assert that there is not enough evidence that the association between gabapentin and pain reduction is significant. This assertion is also true if the confidence interval traverses the line with 1 on the vertical axis. This particular example (RR crosses the “1” line) does NOT mean that gabapentin treatment completely lacks “real world” effectiveness. Indeed, in some patients, use of gabapentin may well have a beneficial effect, but there is not enough evidence in a statistical sense to say that this intervention will be effective.
If one just adds up the numbers of people and events from a number of trials (as in the two rows entitled “Total (95% CI)” and “Total events” near the bottom of Figure 4.4 above), reviewers effectively treated it as one big trial. In effect, people in the treatment group of one trial will be compared with people in the control group of another trial. This comparison is not randomized, and it is likely that there will be some differences in the way the trials were carried out.

### 17.1.1 Limitations of Relative Risk Calculations

Clinicians usually prefer to know the risk that their patients will develop a particular disease called the ‘background’ risk of an event before recommending a treatment. In a placebo-controlled trial, the background or baseline risk is the presence of an event in the control group (i.e. the group without the pharmaceutical intervention under study). The expected benefit of treatment could vary considerably as the baseline risk changes (See original Report Annex 4.1). For example, an estimated risk reduction of 50% relative to placebo in a clinical trial might be statistically significant and clinically important for patients at high risk of a particular event, but not be clinically important for low risk patients.

Thus, treatments with very large efficacy benefits may have a small clinical therapeutic effect where the population comprising the potential patients are healthy or have a very low likelihood of having a poor outcome. Thus, for patients with a low probability of an event, or otherwise healthy patients, this reduction in risk might not be sufficient to warrant the toxicity and cost of treatment. In brief, if a particular drug requires 20 patients to be treated for 1 patient to obtain benefit, this means that 19 patients have to be subjected to the costs and side-effects of the drug, with little prospect of therapeutic benefit. Moreover, if patients have a very low probability of actually getting the condition - a background risk of 1% (.01) - there will be many more people that have to be treated for 1 patient to obtain benefit. This means that it is quite possible that the risks associated with the use of the drug would outweigh the potential benefits. This is one criticism of relative measures of treatment effect in clinical trials for the purposes of clinical decision making. Notwithstanding this, the purpose of this present methodology is to provide guidance for policy-makers and those interested in research prioritization, and it is believed this critique of risk measurements for clinical decision-making loses some of its power.

For some treatments and conditions the benefit of a specific treatment, as measured by the relative risk, remains roughly constant over patient populations at varying baseline risk. In these cases, relative measures appear attractive since a single estimate of treatment effect can be provided for a broad class of patient risks. The relative risk of a given treatment, such as statins for the prevention of ischaemic heart disease, tends to be independent of the risk of the patients being treated. This makes it a good measure to use when combining the results of different trials in a Cochrane systematic analysis.

### 17.2 Calculating Clinical Efficacy with Odds Ratios

As mentioned, in the present report only weighted values of the relative risk (RR) and the weighted odds ratio (OR) will be used as another clinical efficacy metric where these are given in the Cochrane Review. Indeed, many clinical efficacy data in the CDSR are calculated as odds ratios. These measure the same outcomes as relative risk, but in a slightly different way. Briefly, odds ratio and relative risk constitute parallel statistical metrics for measuring
frequency and ratios of frequency. Their relationship might be compared to the use of different scales such as Fahrenheit and Centigrade to report absolute values and relationships between different temperatures.

Risk is a proportion or percentage of an entire population having a given characteristic or outcome. Odds is a ratio of those with and those without the characteristic or outcome within the population. It is important to know that, as risk falls below 20%, odds ratios and relative risk measures become more and more similar and virtually identical below a risk of 10%. Put another way, if the baseline risk (the risk of adverse outcome in the control group) is low (less than 20%), the difference between RR and OR is unlikely to be important. Therefore, the two measures may be used interchangeably for interpreting results. Moreover if the OR is near 1.0, the difference between RR and OR is also unlikely to be important. If the baseline risk is high (over 30%) and the OR is not close to 1.0, the RR and OR might yield results sufficient to bias conclusions.

As the CDSR routinely publishes data with odds ratios, it was initially assumed that the two conditions (high baseline risk, OR not close to 1) are not present. Since it cannot be quantified what an OR “close to 1.0” really means, it was therefore chosen use the weighted OR to supplement clinical efficacy metrics using the RR.

17.3 Calculating Clinical Efficacy with Standardized Mean Difference

Some studies neither report odds ratios nor relative risk ratios as their outcome measure. This is mainly the case for studies on diseases whose severity is measured with a continuous scale. Examples are depression, which is often assessed with a Hamilton Depression scale or similar instrument; osteoarthritis is measured with a pain scale; or dementia is measured with a scale to assess cognitive functioning.

Instead of risk or odds ratios, these studies often report mean difference scores, calculating the mean (average) score on the scale before an and after an intervention and subtracting these averages from each other to quantify the difference. If all studies would use the same unit of measurement (for example average weight in kilograms), it would be easy to compare these mean difference scores, but this is often not the case. If one study uses kilograms as a unit and another study uses pounds, the results are not comparable. One solution to this problem would be an inter-conversion factor to convert one unit into the other. However, if this factor is not known, such a conversion is not possible. A way around this is to compare standardized mean differences, rather than actual means. This can be done with the standardized mean difference.

The standardized mean difference is calculated as the difference between the means, divided by the average standard deviation for the groups. The number represents how many standard deviations the groups differ by. The significance of the standardized mean difference is that its value does not depend on the measurement scale. For example, one could consider a trial evaluating an intervention to increase birth weight. If two studies are found, but one uses kilograms and the other uses ounces, the outcomes can still be compared by using this measure. If the mean birth weights in the intervention and control groups were 2700 gram and 2600 gram with an average standard deviation of 500 gram, the standardized mean difference would be 0.2 ((2799-2600)/500). If the other study would find the exact same
weight in ounces (oz.), the means would be 95 oz. and 92 oz., with an average standard deviation of 15 oz. The standardized mean difference would be the same at 0.2 \((95 - 92)/15\). Because meta-analyses combine different trials, this measure is often used to combine results. Reviewers used the measure in extracting data from Cochrane reviews in addition to results given by relative risk ratios and odds ratios.

18 Modifying the OR and RR Data for Graphical Presentation

The risk metric (OR, RR) is greater than one for outcomes in which the events in the treatment group are more than those in the control group (see Figure 4.4 and RR calculation). However, the metric is less than one for outcomes in which the events in the treatment group are less than those in the control group. This is to be expected for certain curative events (i.e. fewer patients in the treatment group died, fewer patients in the treatment group had a reoccurrence of the condition, or had fewer trips to the emergency room, and so on).

As noted in Figure 4.3 a sizeable minority of analyses showed risk ratios (RR, OR) greater than one which means that more patients in the treatment group experienced an outcome event. Specifically, more patients in the gabapentin group had some measure of pain reduction. In these cases, the outcome measurements were carefully reviewed to confirm if, in fact, a benefit did exist. In all cases where the metric was greater than one and no adverse event occurred, the outcomes measured were consistent with a beneficial effect of treatment (i.e. more patients in the treatment group responded to treatment, reported positive results, fewer deaths, fewer visits to the hospital, and so on).

In the convention that the Background analyses have used for purposes of graphing the data, beneficial interventions that are better than placebo have relative ratios greater than one, as shown in Figure 4.4. Interventions less beneficial than placebo have ratios less than one. For policy analyses, it is better to provide graphical outputs in which a risk ratio greater than one are reserved for all beneficial interventions, as people intuitively give higher numbers a better ranking than lower numbers. Therefore, according to the convention adopted in this report, beneficial risk ratios favouring the intervention were recalculated to give an RR or OR of greater than one.

This was done in the following way:

a. For all beneficial outcomes whose ratio was originally less than one, the ratio was inverted to create a true mirror image. Thus, the upper confidence interval (CI) of a metric below the 1 line of equivalence is transformed into the lower CI of its mirror image. Simply taking the reciprocal of, for instance, a pooled RR of 0.86 (CI: 0.55, 1.4) has 14% beneficial effect would be recalculated as an RR of 1.16 (1/0.86) for an average value of therapeutic efficacy of 1.16 or a 16% beneficial effect. Although the numbers are similar (14% and 16%), they are not identical. Further, taking the reciprocal of the confidence intervals would also artificially expand them and possibly distort the asymmetries. Simply inverting the original 95% confidence intervals would have yielded an upper CI of 1.81 (1/0.55) and lower CI of 0.71 (1/1.4), which is a true distortion.
In order to provide a true inverted mirror image reviewers used a simple Excel® formula that calculated the distance between the risk metric, the respective upper and lower confidence interval and the line of equivalence (metric =1.0), and then recalculated the ‘inverted’ distances. What is of most concern in the few inversions that were done was whether an upper confidence interval that traversed the 1.0 line of equivalency was transformed into a lower CI that no longer traversed the 1.0 line, and vice versa. That is, the concern was transforming an outcome that was ambiguous with regard to whether or not the medicine was better than placebo into an outcome that was no longer ambiguous. The Excel® formula to create a true mirror image avoids this.

b. For all beneficial outcomes whose RR was originally greater than one, reviewers left the RR, OR values, and the confidence intervals untransformed (see Figure 4.3).

Figures were created using the Excel graphical program for stock prices, which requires three values in this order: high, low, and close. Values are therefore in sequence: upper confidence interval, lower confidence interval, and efficacy metric (i.e. weighted odds ratio or weighted relative risk ratio). These three numbers are used to create the figures, and an example of medicines to treat myocardial infarction (MI) is found below in Figure 4.4.

The black squares are the inverted efficacy OR and/or RR metric (where appropriately inverted), and this is recorded on the Y axis with their respective 95% confidence intervals and the horizontal line represents the line of equivalency where the metric is 1.0. Based on this convention, points below the 1.0 line mean that the treatment was less effective than placebo. The horizontal axis lists the 14 different meta-analyses of various treatments for MI.

There are two separate Cochrane Review and one BMJ Clinical Evidence summaries, separated by the vertical coloured lines. Within a given review, each metric represents the results of a meta-analysis of the same condition treated by a medicine or group of medicines as listed on the X axis, but will often represent different outcome measurements (i.e. death, time to recurrence, length of hospital stay, and the like) as well as possibly different dosages.

Where the square box is above the horizontal "1" line but the confidence interval line crosses below the horizontal "1" line this means that the benefits shown by the various clinical trials for the particular intervention might have occurred by chance alone. Therefore, in the statistical sense, the intervention has not been shown to have an unequivocal benefit. Figure 4.9.1, shows that in many trials the intervention was more clinically efficacious than the placebo as the lower boundary of the confidence interval is above the horizontal “1” line.
Figure 4.4: Sample graph for medicines used to treat myocardial infarctions

*: RR or Odds ratio > 1 favours the intervention (less mortality or myocardial infarction)

19 Limitations of Cochrane Reviews

There are several limitations of CDSR. Cochrane reviews are necessarily retrospective. Not all medicines are reviewed because many have been used for years (antibacterials, such as older TB drugs, some antiretrovirals, insulins, and older antimalarials), and it would now be unethical to conduct a placebo controlled trial. Of the medicines that are reviewed, some are off patent, some are being used off label, and some have not yet even received regulatory approval. Thus, the repertoire of interventions for a particular condition using Cochrane may or may not be the most up to date.

Thus, these meta-analyses alone cannot tell us about pharmaceutical gaps as more recent information is needed. They can only provide a preliminary screen of potential pharmaceutical gaps. The analysis must be updated to provide information on drug pipelines and drugs in clinical trials. Furthermore, Phase III placebo-controlled trials are usually short, despite the fact that important decisions about prioritization may ultimately be based on evidence of long term costs and effectiveness measures. 36

Although most trials data in the CDSR are placebo controlled trials, funding decisions should be based on comparison of new interventions with current practice, not with
placebos. From a public health viewpoint, what is important is not whether one intervention works better than no intervention, but whether it works better than alternative interventions. This is an important, albeit understandable, omission as head-to-head comparisons are both expensive and time consuming (see Background Chapter 8.4).

In addition, it is well known that trials with negative outcomes are underreported in the literature so publications describing pharmacological interventions that do not work is difficult to find. This publication bias is a difficult problem to overcome. Indeed, conducting a Cochrane meta-analysis does not overcome problems that were inherent in the design and execution of the primary studies.

Perhaps most importantly, significant outcomes from a public health viewpoint may not always be evaluated using RCTs. For instance, it is difficult for RCTs to evaluate unpredictable things such as adverse events induced by drug use. Characteristics of these analysis and discussion, together with poor quality reporting, may perhaps justify a systematic review focused on specific adverse outcomes and inclusion of observational studies.

20 Data sources for R&D funding for “neglected tropical diseases, diarrhoea, pneumonia, tuberculosis, HIV/AIDS and malaria”: the G-FINDER

Donors interested in funding R&D of products for neglected diseases must currently make substantial investment decisions in the absence of accurate data regarding funding flows, gaps and duplications. Information that is available is often out of date, patchy and unreliable or cannot be compared across surveys due to different accounting and reporting methodologies. In some areas there is an almost total lack of information.

The goal of the G-FINDER survey is to help funders to better target their investments into neglected disease product R&D (see http://g-finder.policycures.org/gfinder_report). G-FINDER tracks global investment annually in this area. It is hoped that by providing funders with better information, the G-FINDER survey will stimulate increased efficiency and investment into neglected disease product R&D. The G-FINDER survey includes 31 neglected diseases, and the pharmaceutical tools used to prevent, control and treat these, including medicines, preventive and therapeutic vaccines, diagnostics, microbicides and vector control products. The survey encompasses R&D funding for these products from basic research through clinical trials.

21 Conclusions

The methodology described above serves as the basis to determine priorities for a public health-based research agenda for the EU. This is accomplished using in-depth analyses to decide which pharmaceutical gaps should be studied further. As discussed in Chapter 3, the prioritization scheme employs several different conceptual frameworks. The "ethical/moral"
conceptual framework for prioritization is summarized in Chapter 3. The "evidence based" framework is described in this chapter and combines burden of disease and risk factor analysis with clinical efficacy measurements, as updated by informed judgments about future clinical and epidemiological scenarios plus in-depth Background reviews of recent science, drug development, and market analyses for multiple disease areas.

References


Contributors to the Cochrane Collaboration and the Campbell Collaboration. Evidence from systematic reviews of research relevant to implementing the ‘wider public health’ agenda York: NHS Centre for Reviews and Dissemination, 2001. www.york.ac.uk/inst/crd/wph.htm (accessed 17 July 2001)


