Generalizability of Health Economic Evaluations: Methods for Multinational Patient-Level Studies

Thesis

By Anna Filonenko, MSc

Presented to the Faculty of Health Sciences Bielefeld University

in partial fulfillment of the requirements for the degree of Doctor of Public Health

1st Advisor: Prof. Dr. Wolfgang Greiner, AG 5 Gesundheitsökonomie und Gesundheitsmanagement.

2nd Advisor: Prof. Dr. Ralf E. Ulrich, AG 8 Demografie und Gesundheit/Institut für Bevölkerungs- und Gesundheitsforschung.

Bielefeld, 01.06.2010

Printed on DIN ISO 9706 paper

Abstract

Background

Cost-effectiveness outcomes collected in multinational health economic studies may vary across countries and geographic areas due to numerous clinical and socio-economic factors. The between-country variability poses the question of generalizability of the cost-effectiveness estimates, i.e. applicability of the trial-wide results for the decision-making in particular jurisdiction.

Objectives

The purpose of this research is to illustrate how selected methods can improve generalizability of the results of patient-level health economic studies by a) exploring between-country variability in incremental costs, effectiveness and resource use, b) calculation of trial-wide and country-level incremental cost, effectiveness and resource use while accounting for patient- and country-level covariates, and c) assessing the potential of covariates to predict cost and effectiveness estimates for the settings outside the study.

Methods

Review of published health-economic evidence and international HTA guidelines was conducted to evaluate applied or recommended methods and analytical strategies to improve generalizability of the health economic outcomes. In the case study, qualitative and quantitative homogeneity test by Simon and Gail is used to explore between-country heterogeneity of the treatment effects measured by incremental costs, effectiveness, resource use (length of hospitalization) and incremental net monetary benefit. Hierarchical modelling is applied to calculate trial-wide and country-level estimates, while accounting for clustering and incorporating country- and patient-level covariates.

Results

Simon and Gail test indicated qualitative homogeneity for incremental effectiveness, costs, length of hospitalization and net monetary benefit between treatment and control arms. Trial-wide and country-level mean incremental costs and effectiveness were estimated using hierarchical models with and without covariates. Results of hierarchical modelling suggested, that new treatment was more efficacious, saved costs and resource use in majority of the countries.

Conclusions

Homogeneity test and hierarchical models are complementary methods to explore heterogeneity and to estimate trial-wide and country-level parameters for cost-effectiveness analysis. Hierarchical models allow for country-level estimates and adjustment for covariates. In the case study the patient-level covariates showed effect on incremental cost and effectiveness; country-level covariates had very small impact on estimates. The use of country-level covariates as predictors for incremental cost and resource use to improve generalizability of health economic studies beyond the study setting merits further investigation.

Preface and Acknowledgments

I would like to thank Prof. Wolfgang Greiner for his guidance and support of this work. I am also very grateful to many current and previous colleagues in Schering AG and Bayer Schering Pharma AG for supporting advices. Special thanks goes to Prof. Mark Glickman for his comments and suggestions for Bayesian analysis. Very special thanks goes to Dr. Kerstin Uhl-Hochgraeber for invaluable help, motivation, pragmatism and sense of humor. Above all, my husband and son keep encouraging all my endeavors (in very different, but efficient ways) including this research, which I would have probably not pursued and definitely not accomplished without them.

Table of Contents

List of Abbreviations

AIDS	Acquired Immune Deficiency Virus
AMCP	Academy of Managed Care Pharmacy (US)
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
BE	Belgium
BP	British Pound
BUGS	Bayesian Inference Using Gibbs Sampling (software)
CADTH	Canadian Agency for Drugs and Technologies in Health
CAP	Community-acquired pneumonia
CEAC	Cost-effectiveness acceptability curve
CH	Switzerland
CI	Confidence interval
CNTR	Country
CR	Credibility interval
CRF	Case report form
DE	Germany
DIC	Deviance Information Criteria
DM	Deutsche Mark
DRG	Diagnosis Related Group
e.g.	For example
EARSS	European Antimicrobial Resistance Surveillance System

ES	Spain
et al.	And others
etc.	And so forth
EUR	Euro
FDA	Food and Drug Administration
FF	French Franc
FR	France
GDP	Gross Domestic Product
GKV	Gesetzliche Krankenversicherung
GOÄ	Gebührenordnung für Ärzte (Germany)
GR	Greece
HIV	Human Immunodeficiency Virus
HM	Hierarchical models/modelling
HRQoL	Health-related Quality of Life
HSE	Health Services Executive (Ireland)
HTA	Health Technology Assessment
i.e.	That is
ICC	Intraclass correlation coefficient
ICER	Incremental cost-effectiveness ratio
ICU	Intensive care unit
IL	Israel
INFARMED	National Authority of Medicines and Health Products (Portugal)
INMB	Incremental net monetary benefit
IPHA	Irish Pharmaceutical Healthcare Association

IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ISPOR	International Society For Pharmacoeconomics and Outcomes Research
ITT	Intention-to-treat population
IV	Intravenous (administration)
KCE	Belgian Health Care Knowledge Centre
LOS	Length of (hospital) stay
MCMC	Markov Chain Monte Carlo
ML	Maximum likelihood
MS	Multiple Sclerosis
NA	Not applicable
NGAP	National Fee Schedule (France)
NHS	National Health Services
NICE	National Institute For Clinical Excellence
NIE	New Institutional Economics
NMB	Net monetary benefit
NSAID	Non-steroidal anti-inflammatory drugs
OECD	Organization of Economic Co-operation and Development
PHARMAC	Pharmaceutical Management Agency of New Zealand
PMSI	Programme de Medication des Systemes d'Information (France)
PO	Per oral (administration)
QALY	Quality adjusted life years
RU	Russia
SD	Standard deviation
SFr	Swiss Franc

SHI	Statutory health insurance
SMC	Scottish Medicines Consortium
SOIKOS	Base de datos de costes sanitarios (Spain)
TIPSD	Tarif Intermisisterel des Prestations Sanitaires (France)
TNF	Tumor necrosis factor
TOC	Test of cure
UCANSS	l'Union des Caisses Nationales de Securite Sociale (France)
UK	United Kingdom
WBC	White blood cells
WHO	World Health Organization
WinBUGS	BUGs (software) for Windows
ZA	South Africa

List of Tables

List of Figures

1 Introduction and Research Purpose

Data collection in multinational trials

Cost-effectiveness analysis for informing health policy decision making is often conducted using patient-level data from multinational clinical trials. There are two main advantages of conducting the trials in several jurisdictions or countries. The first practical advantage is potentially fast recruitment of the study patients. The second advantage relates to the diversity of recruited population coming from different locations with different clinical and socio-economic background. The latter could potentially improve generalizability of health economic outcomes beyond the study setting. However, variation in multiple health economic parameters across the jurisdictions raises the question about the generalizability of the study results.

Definition of generalizability

Applicability of the health economic evidence collected in multinational trails for the decision-making in particular jurisdiction is a complex issue. Clinical and economic parameters collected in different settings vary within and between countries due to numerous clinical and economic factors. As a consequence, trial-wide cost-effectiveness estimates may not be directly applicable for the decision-making for reimbursement of a therapy in a particular country (Sculpher et al., 2004; Drummond & Pang, 2001; Boulenger et al., 2005; Mason & Mason, 2006; Li, 2007). This issue is discussed in health economic literature using different terminology, such as transferability, portability, extrapolation, external validity and others. For the purpose of this research we will use the definition suggested by Sculpher and colleagues (2004), who defined generalizability as "the extent to which the results of a study based on measurement in a particular patient population and/or a specific context hold true for another population and/or in a different context" (Sculpher et al., 2004, p. ix). Further on, for the analysis of patient-level studies we will refer to "internal" generalizability in considering the validity of the results for the countries or settings included into particular study, and to "external" generalizability in discussion about applicability or adjustment of the estimated outcomes to the external settings not participated in the study.

Current methodological approach and need for further research

Assessment of the generalizability of the results is not yet established part of the analysis of multinational health economic trials. It is generally assumed, that clinical efficacy of a therapy should be similar across different countries. Homogeneity of relative clinical effects is usually tested before the data from multinational trails is pooled. However, statistically significant differences in treatment effects in different countries are rarely observed or may go undetected due to small sample size (Cook et al., 2003). The variation in resource use and cost between the countries posses a challenge for cost-effectiveness analysis, as it may prevent from pooling of economic parameters and not allowing for robust sub-group analysis with country-specific data due to small sample size. The application of the unit prices from one country to the pooled resource use across all participated countries was often done in early health economic studies as a way to overcome small sample size issue in individual countries. The appropriateness of aggregating the economic data across the jurisdictions is rarely investigated.

The acknowledgement of methodological issues and growing demand for qualitative data for jurisdiction-specific decision-making with increasing volume of multinational health economic studies, has rapidly evolved the development of analytical methods to improve generalizability. Task Force of International Society for Pharmacoeconomics and Outcomes Research (ISPOR) has been set up and worked out a guideline to address this issue (Drummond et al., 2009). With regard to patient-level trials, homogeneity test and hierarchical models have made their way in the literature. These methods are also suggested by ISPOR guideline as possible strategies to explore heterogeneity and account for variability and hierarchical data structure (Drummond et al., 2009). Broad application of these methods in studies conducted in different locations and clinical indications will help to understand and quantify between-country heterogeneity, improve the generalizability of the results and, thus, increase the volume of available evidence.

Research purpose and objectives

The overall aim of the research is to explore analytical methods to improve generalizability of multinational health economic evaluations. The purpose of the thesis can be specified in terms of the following research objectives:

1) To review available methods to assess heterogeneity and to improve generalizability of health economic evaluations conducted in multinational settings.

2) To evaluate applicability of selected methods to the patient-level data in a case study investigating between-country variability of cost-effectiveness parameters, heterogeneity factors explaining between-country heterogeneity and to calculate country-level and trial-wide cost-effectiveness estimates.

Analytical approach of the case study is essentially the following. We first explore between-country variability and hypothesize major heterogeneity factors. We then calculate country-level and trial-wide estimates for study settings (internal generalizability). Finally, based on heterogeneity factors and country-level estimates, we explore possibility to generalize results to non-study settings (external generalizability).

Thesis structure

This thesis is structured as following. Chapter 2 discusses Public Health relevance of the research subject, describes theoretical link between Public Health and Economics, and outlines the theoretical framework for Health Technology Assessment (HTA) and health economic evaluations. Chapter 3 summarizes current analytical approaches to improve generalizability described in published evidence and methods used in the case study analysis. Specifically, we used homogeneity test to explore and assess variability/country-by-treatment interaction for incremental effectiveness, costs, length of hospitalization and incremental net monetary benefit with implications for internal generalizability, i.e. applicability of the results for the decision-making in the countries participated in the study. Hierarchical models were used to calculate country-specific estimates with and without covariates, to assess covariates effect on cost and effectiveness estimates and possibility to generalize results of the study to non-study settings on hand of explored covariates, i.e. external generalizability. Chapter 4 describes dataset, costing methodology and results of a case study. Chapter 5 discusses findings, limitations of the methods and analysis, and suggests possible extensions for the future research. Finally, chapter 6 provides top summary of the results and conclusions.

2 Theoretical Perspectives on Public Health and Health Economic Evaluations

Methods to improve generalizability are essentially analytical tools for application in the analysis of multinational studies. Health economic studies, being part of HTA, are deemed to inform political decisions. Health policy and so the informing evidence should be jurisdiction-specific. Here, the generalizability methods aid to customize available data to the specific decision context. Development and application of the analytical tools, such as generalizability methods, usually require underlying theoretical considerations.

Theoretical foundations of health policy and research methods for health economic evaluations are rooted in Public Health and Economic realms, what is the subject of this chapter. We review theoretical and instrumental concepts related to Public Health and policy to understand theoretical implications for generalizability research and to establish the relevance of the generalizability issue to Public Health context. The discussion will start with normative theoretical perspectives of Public Health, Neoclassical Economics, Non-welfarism and New Institutional Economics. We then move towards instrumental concepts such as health policy, HTA and health economic evaluations, concluding with arguments about relevance of generalizability methods for Public Health, policy and research.

2.1 Normative Theory and Instrumental Concepts

Need for normative analysis

Methods to improve generalizability of health economic evaluations are in essence analytical instruments. Why then does normative theoretical analysis matter? Welfare economic theory suggests, that characteristics of health care market may impede optimal allocation of resources and public intervention should correct free market

failure. Indeed, great proportion of health care expenditures in the developed countries is financed from the public budget either explicitly or through taxation (OECD, 2005). Non-market institutions and political instruments play essential role in regulation and allocation of resources. Therefore, normative analysis should guide both policy formulation and empirical research.

There are many possible ways and instruments to implement political decisions. If the role for the invisible hand working through markets is in practice quite circumscribed in health sector, how do we know, what health policy and institutional designs will produce an efficient and equitable allocation of health care resources? Indeed, it is a difficult task to demonstrate that policy failed, but it is even more complex to design the one which will succeed. Normative economics is precisely about the attempt to determine the "ideal" or "desired" state, resource allocation and political measures to achieve it. Market failures, that pervade health care sector, create an important role both for non-market institutional arrangements (i.e. the "visible hand") and for normative economic analysis to help design the "good" policies (Hurley, 2000).

Different methods for assessment of outcomes such as health benefits, preferences and utility could produce different estimates and lead to contrary political decisions and implications. Examination of theoretical foundations can provide a rationale for selecting appropriate analytical methods. The validity of the measured outcomes using specific methods should be verified by underlying theory. More importantly, recourse to fundamental theoretical considerations will help to maintain continuity and integrity in political decision-making (Hall et al., 2006).

Link between Public Health, economic theories and generalizability methods

Understanding the relationship between theoretical and instrumental concepts related to health policy will help to identify implications for generalizability research. Figure 1 provides a graphical presentation of the logical links between health economic theories and Public Health instruments. Generalizability methods are anchored within HTA domain. In our theoretical quest we first step back and start with neoclassical economic theory, followed by newer health economic theories navigating toward health policy, HTA and health economic evaluations as instrumental concepts.

The logical connection between individual constructs is essentially the following:

1) Public Health and Economics.

- Public Health and Economics are linked empirically: population health is a national resources and directly influence economic performance and activities.
- Neoclassical theory advocates public intervention (health policy) to correct free market failure in health care.

2) Beyond Neoclassical theory toward Public Health paradigm.

- The assumptions of neoclassical theory (utility maximization, welfarism) have been challenged for the application to health care.
- Alternative theories, such as non-welfarism, extra-welfarism, communitarianism and others have contributed to convergence of Public Health and Economic paradigms.

3) New Institutional Economics (NIE) to guide health policy.

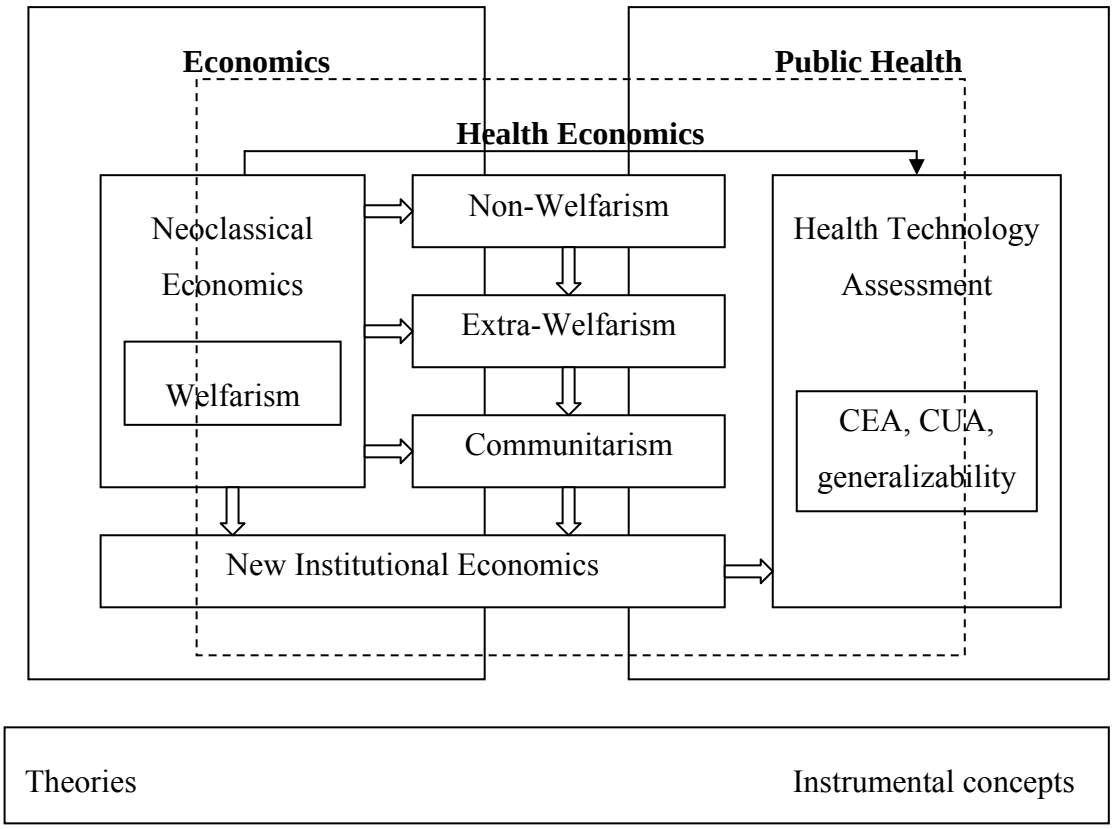
- Neoclassical framework is of limited use to study health policy, as it is essentially institution free; New Institutional Economics provides an alternative theoretical foundation to study incentives, efficiency and policy in health care.
- New Institutional Economics can aid to explain geographic heterogeneity in inputs and outputs of health care market as well as to study and evaluate outcomes of different political instruments.

4) HTA and health economic studies to inform health policy.

- Health policy should be informed by jurisdiction-specific evidence in order to improve health and reduce inequalities.
- Generalizability methods aid to customize available limited evidence to reflect jurisdiction-specific context.

We elaborate the above arguments in more details in the following sections.

Figure 1: Relationship Between Public Health, Economic Theories and Instruments



Source: compiled by the author

2.2 Public Health and Neoclassical Economics

Current Public Health concerns

In 1952 WHO defined Public Health as a mean of promoting health and preventing diseases through organized efforts of society (WHO, 1952). Despite the differences in health care systems across the countries, essentially any national health policy should aim to achieve Public Health goal – improvement of community health.

Although the question about financing of the health care systems seems to be unavoidable, there is a loud criticism of the involvement of economic consideration about access to medical technologies in the political decision making process. The opponents warn of the "economization" and "medicalization" of the health care systems, what leads to disproportional increase in drug expenditures and distract political focus and resources from the public health measures, e.g. prevention

programmes. Indeed, many health historians agree, that medical therapies played only negligible role in the historical decline of the population's mortality rates. Effective medicine is fairly recent phenomena. Production of health care services in the magnitude affecting national health was only possible in the twentieth century. It is clear, that public health measures and rising living standards attribute to the largest share of credits in the community health improvement (Folland et al., 2006).

Technological progress contributes significantly to the increasing expenditures in the health care sector. The introduction of the economic mechanisms to control the costs with e.g. "physician drug budgets" and "patients co-payments" does not always find support of clinical community. Potentially these mechanisms could place a significant financial burden to the disadvantaged stratum of the society and increase inequality (Rosenbrock, 2001; Kühn & Rosenbrock, 1994; Lützenkirchen, 2000).

Facing the above concerns, we first investigate the complex relationship between national economy and global (community) health and will seek to understand, in general, why Health Economics should have it place in Public Health realm, and, in particularly, what is the theoretical rationale for health policy.

Wealth and Health of Nations

This and following sections explore the macro-level relationship between economic performance and population health. Under "macro-economy" and "population health" we understand economy and community health on the national level, although the application to the worldwide regional levels (e.g. European Union, North America) is also possible. Presented evidence suggests the direct or indirect dependence between national economy and community health. National economy benefits from community health improvement and vice versa. This is a rational for not to study Public Health measures explicitly apart from economical considerations.

The dependency and interactions between national economy, health provision and outcomes is a complex issue. The majority of these interactions can be separated into two groups. The first group focuses on impact of health on macro-economic activities, such as economic growth, labor supply and consumption. The second group considers the reverse effect of economic development on health outcomes and national health care system (Rugger et al., 2001).

Health influences Wealth

Population health status influences economic growth and development directly in e.g. provisioning of labor supply and labor productivity, level of consume and preferences, saving rates, utilization of health services.

Health and health care are generally treated as economic goods in both developing and industrialized countries with the respective attributes: price, supply and demand. From macroeconomic perspective, health sector plays major role in global economic activity with ca. 10% of total global product in 1998 and has significant potential to affect economic development and growth (Rugger et al., 2001). In 2003, countries of Organization of Economic Cooperation and Development (OECD) dedicated ca. 8.8% of their gross domestic product (GDP) to health spending comparing to 7.1% in 1990. The proportion of GDP spent on health care varies from 15% in the United States, followed by Switzerland and Germany with 11.5% and 11.1% to less than 6% in the Slovak Republic and Korea (OECD, 2005).

Considering health as a human capital (elaborated by Becker, 1964), high mortality, early retirement and reduced labour productivity can ultimately contribute to reduction of income at individual and national level (number of studies available, see e.g. Thomas & Strauss, 1998).

Poor health status and disability have economic consequences on individual and national levels, as the production of health is financed from GDP through public or private channels. This is of relevance for both wealthy and low income economies. For example, a study of the impact of HIV/AIDS on national economy of India estimated annual costs of HIV/AIDS roughly at 1% of the GDP (Anand et al., 1999). The industrialized nations are facing the challenge of financing the treatment and care of the increasing number of patients with chronic diseases (Rosenbrock, 2001).

Wealth influences Health

The reverse linkage reflects the influence of economic development on health and health care production. In fact, life expectancy at birth is positively linked to the level of national income in OECD countries (Rugger et al., 2001). However, life expectancy varies substantially across countries with similar level of economic development. Based on level of GDP per capita, projected life expectancy in Hungary and US is higher than

the real value and vice versa in Japan and Spain, where real life expectancy is higher than predicted outcome (OECD, 2005).

The trend of rising health care expenditures is common across industrial and developing countries. Several causes of this trend are being discussed. Increasing drug costs are of particular concern. From 1997 majority of OECD countries observed above 5% annual increase of drug costs, which outpaced the growth of overall health care expenditures. Although the trend of rising costs is global, there are substantial differences in resource distribution within various health care systems. This variation may be associated with different non-market institutions, structures and political decisions, which have impact on access, reimbursement and implementation of new medical technologies (OECD, 2005).

The differences in health expenditures and health outcomes suggest significant variation in efficiency in the health care systems, with which they "produce" health. Some countries "produce" more health for less money; others achieve similar outcomes by higher level of expenditures (Rugger et al., 2001). Rising expenditures itself do not indicate inefficiency of resource allocation or production of health services. Development of new technologies in the areas where no treatment where available and increased willingness to pay for maintenance of good health or health gain could justify dedication of additional resources (Schwermann et al., 2003; Schulenburg & Greiner, 2000). Nevertheless, given the changes in health care production, in the demand for new technologies and in the underlying demographic and epidemiological factors, the efficient allocation of resources is necessary to maximize health output.

As discussed above, the relationship between health and economy is complex (see Figure 2). Ultimately, wealth is health and health is wealth (Brouwer et al., 2006; WHO, 1986). Despite the critics of "economization" of health provision, the market for health care production exists and is growing. For many industries the "invisible hand" of free-market leads to the efficient allocation of resources. If it so, could we gain potentially more health without regulating the market with political instruments? Following sections investigate the applicability of the competitive (free) market conditions to the health care.

Efficiency of "invisible hand"

The production and distribution of health depends on the functioning of health care market. Economic theory suggests that for many goods and services the mechanism of competitive market will force efficient allocation of resources – the "invisible hand" solution (The First Fundamental Theorem of Welfare Economics). According to the economist Wilfred Pareto, the efficient allocation will be the optimal outcome for the society, as no one will be able to improve his lot without "hurting" someone else. This implies, that Pareto efficient outcome can be achieved in a competitive market (Second Fundamental Theorem of Welfare Economics) (Mankiw, 2008).

In most cases, health care is a mean to obtain or restore particular health status. It is an intermediate product. The concept of efficiency of health care production considers the ratio between input factors (resource use, investments and labour) and intermediate or end products (prevented cases, successfully treated patients, QALYs). Efficient production means that the outcome is maximized at a given resource endowment. The analysis focused on intermediate parameters (e.g. duration of hospitalization) without considering health endpoints, may misinform or provide partial evidence for health policy decisions (Drummond et al., 2005). From Public Health perspective, health outcomes and not intermediate products should be maximized.

In fact, health care provision is highly regulated in the majority of countries. But if "invisible hand" solution does not need a hypothetical decision-maker heavily armed with health economic data, why do health policy and institutions matter?

Pre-requisites of free market economy

In 1963 economist Kenneth Arrow demonstrated necessity of public intervention in his milestone article "Uncertainty and the Welfare Economics of Medical Care" (Arrow, 1963). Arrow explored the "uncertainty in the incidence of disease and the efficacy of treatment" which contributed to the inefficient allocation of the resources in the competitive market. Adequate non-market institutions (health economic decision-makers) will correct and compensate the free-market failure. Since 1963 the changes in global economy and health care market are being accelerated by technological progress. However, Arrow's publication stood the test of time. It provides the insight into the root of problems of the health care market and is broadly cited in Health Economics and other disciplines today.

The major criticism of the applicability of the "invisible hand" solution for the health care market concerns the vital condition of the free market competition (or market ability to obtain competitive framework). Perfectly competitive market must have no entry or exit restrictions for the market players; all participants should be perfectly (symmetrically) informed and trade with a homogenous product; numerous buyers and sellers should not have sufficient power to influence the price. Perfect market does not allow for uncertainty in events, externalities, monopolies or public goods (Folland et al., 2006). The health care market departs from competitive market in a several ways. Selected issues are discussed in the following.

Violations of free market conditions in health

Condition of perfect information is a key assumption for optimality of free-market. The market players should have the same (symmetric) information to be able to assess the value of the product. In health care, however, patients are very often not in the position to evaluate the effectiveness of the treatment. No are the patients able to make a proper choice for the therapy, because of uncertainty in diagnosis. The information about the supply (e.g. treatment) and the demand (patient health/disease condition) is distributed asymmetrically between the pharmaceutical companies, medical personnel and the patients. Associated problem with the patients' lack of information is supplier induced demand, what can lead to the unnecessary use of resources (e.g. inappropriate drug-intake). Thus, the actions to enhance information of all market players could lead to improved allocation. Nevertheless, the fulfillment of the condition of symmetric information is rarely possible (Arrow, 1963; Witter et al., 2000; Folland et al., 2006).

The uncertain nature of health status, diagnosis and treatment outcomes violates another condition of free-market. The occurrence of disease and it severeness is often difficult to predict. Individuals are not able to calculate the risk and budget for health care, as they do for regular goods. To pool the risk and distribute the costs across population requires an insurance system. Insurance tends to change the prices of care to the insured population. Insurance companies or government (being in monopoly position) can often negotiate the prices with the suppliers, removing herewith the price setting mechanism from the market completely (Arrow, 1963; Witter et al., 2000; Folland et al., 2006).

Externalities occur in health care markets, where either production costs or utility of particular goods are not limited to the actors engaged into the transaction of this good. A positive externality may occur to the patient's partner, when the first has been treated successfully. Another example is vaccination, which provides protection for immunized person and those who might come into contact. Negative externality can occur to the family members of the patients suffering from severe chronic diseases, where the burden of the family care-givers can not always be adequately measured and reflected in the therapy price. When the externality occur, the price of the product does not reflect it benefits, what can lead to over- or under- consumption comparing to the social optimum (Witter et al., 2000; Folland et al., 2006)

Social justice is inadequately addressed within free-market economy

Ethical and equity dimensions go far beyond economic allocative efficiency, and therefore free market equilibrium should not necessarily reflect socially desired outcome. Besides, equity and social justice are complex issues and do not have universal understanding (Schreyögg, 2004). Following the "difference principle" from one of the modern concepts "A Theory of Justice" (1971) by John Rawls: "Social and economic inequalities are to be arranged so that…they are to be of the greatest benefit to the least-advantaged members of society …" (Rawls, 1971, p. 303). To pursue this principle under the condition of inverse gradient relationship between socioeconomic status and health level in the society, a redistributive mechanism should be introduced in the economy (for inverse gradient see Rosenbrock, 2001). Health policy should focus on enhancing equity in financing and delivery of the health care services, what is challenging for both developing and developed countries. Current evidence suggests that the political agenda in industrial countries is subject to controversial discussions about the health care objectives, financial arrangements and distributional considerations of costs and outcomes (Wagstaff et al., 1999; Call to Action, 2008).

Many health care systems have introduced the mechanisms to control the costs and reduce overuse of health care through co-payments, drug budgets, medical consultation fees paid by the patients. Studies have shown that these instruments indeed reduce health care use; however, they are "blunt" in achieving their objective. In most cases, they are not able to distinguish between appropriate and inappropriate use by deprived or well-off individuals. The use of "blunt" instruments could have a disproportional

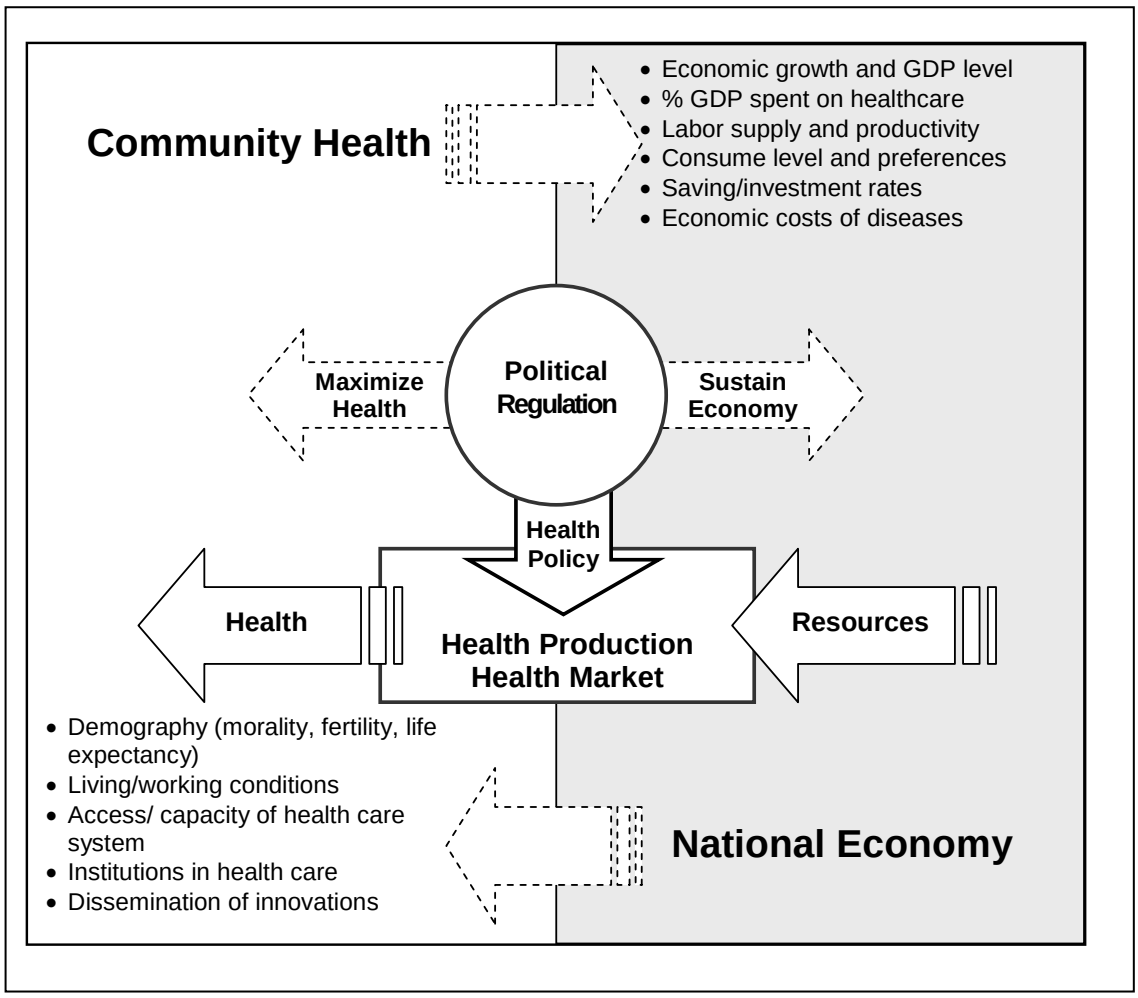
impact on poor and medically indigent (Rugger et al., 2001; Lützenkirchen, 2000). Health policy should be conducted in hand with economic expertise to achieve desired equitable public health outcomes and distributional effects.

Health markets need political regulation

The issues discussed above explain several violations of the competitive market conditions as applied to health care: consume is driven by the expected health status with high uncertainty around treatment effects and health outcomes, positive and negative externalities, and asymmetrical information.

Separately the above aspects exist in other markets, but the interaction and combination of these issues poses a challenge for sound economic analysis and health policy (Folland et al., 2006; Hurley, 2000). We must accept that the real world health care market perfection is not fully attainable, what will inevitably lead to the free-market failure to achieve social optimum. The wealth and health of nations are not only the matter of "invisible hand". The market failure justifies political regulation through non-market institutions (Arrow, 1963; see Figure 2). We are coming back to necessity of a decision maker (or political intervention), who should, where possible and necessary, encourage more competition and let the market mechanism play it effective role and, in situations where the market fails, correct the resulting inefficiencies.

Figure 2: Macro-link Between National Health and Economy



Source: compiled by the author

2.3 Beyond Neoclassical Theory Toward Public Health Paradigm

In the previous section we concluded: (1) from theoretical and empirical perspective, Public Health and Economics should not be studied explicitly apart from each other and (2) there is need of political intervention to correct free market failure.

Health care is different from other economic sectors in the ways it generates market failure. The complex relationship between economic, social and cultural factors requires a multidisciplinary approach to the normative and positive Public Health research and

re-work of established economic, social, ethical, medical paradigms to formulate sound health policy.

The realism and relevance of mainstream neoclassical theory to Public Health realm has been challenged by many health researches. Dissatisfaction with this conventional model among economists encouraged the development of alternative frameworks in Health Economics known as “non-welfarism” by Sen, “extra-welfarism” developed by Culyer and “communitarianism” elaborated by Mooney. There is no unanimous agreement about the appropriateness and relevance of the newer concepts to the Public Health realm. They share some common features as well as several differences, which are outlined in the next section.

Concept of welfarism

Mainstream health economic analysis is essentially founded in neo-classical welfare economics (Hurley, 1998). The framework of welfare theory is based on the following key features:

1) Consistent and rational preferences.

The assumption about rational behavior and consistent choice. “Neo-classical” individuals are able to compare and consistently rank available options and select the one which will maximize their personal utility.

2) Individual autonomy.

The notion that utility and welfare is a private entity. Individual assessment and choice, and not that of a decision-maker or the society, is relevant for preferences, available options and constrains.

3) Consequentialism and welfarism.

The assumption that available options are valued based on their final outcomes and actions are driven by utility maximization. Action is seen in terms of the end product and not as the process. The individuals are expected to take actions driven by the notion of increased utility and welfare. All non-utility terms, incentives (such as altruism) and processes are being ignored (Mannion and Small, 1999).

Social welfare is a function of the utility levels attained by members of the group, whereas the sources of individual utility are often restricted to consumption of goods and services. Ordinal utility theory relaxed assumption of cardinal assessment and comparability of utility function between individuals and suggested Pareto Optimization instead of maximizing the sum of individual utilities (Hurley, 2000). Resource allocation is considered Pareto Optimal (i.e. allocatively efficient) in the case it is impossible to improve individual or group's utility without diminishing or holding utility of the others constant.

Two fundamental theorems of welfare economics have been influential in setting market allocation as the reference standard in normative economic analysis and in justifying major focus on efficiency concerns over distributional equity. The first theorem predicates that the allocation of resources achieved within perfectly competitive market is Pareto optimal. The second theorem states that any Pareto optimal allocation can be achieved through the perfect competition. These theorems provide the rationale within welfare economics for taking a market allocation as the referenced state. The only reason for non-market arrangements is market failure caused by deviation from one or several model assumptions (Hurley, 2000).

Critique of welfarism: foundation of non-welfarism

There are many arguments pointing out the limitations of neo-classical economics. This section cannot comprehensively review all aspects related to health sector. The issues discussed below are mainly based on Sen's non-welfarism framework (Sen, 1977; Sen, 1987). His capability framework addressed critical assumption of traditional neoclassical economics in application to health care. Alternative theories such as extra-welfarism and communitarianism discussed in the next sections, draw on Sen's work.

The distinctiveness of Sen's alternative approach lies in accentuation the impact of individual functioning and capabilities on personal well-being and health. Sen's concept reaches out to Antonovsky's salutogenic Public Health paradigm (1987), who discussed relationship between health, stress and coping as factors of individual health status (Antonovsky, 1979; 1987).

Sen pointed out several limitations of neoclassical assumptions. Consequentialism and welfarism imply consumption and endowments as essential objectives and outcomes of individual actions. Even if the concept of utility can be extended from consumption to

well-being, consequentialism and aggregating utility over individuals remain critical and not fully feasible for application in health care (Sen, 1987). To address neoclassical flaws, Sen emphasizes personal functioning and capability as intrinsic characteristics of individuals. Personal functioning relates to the ability of individual to execute actions and to benefit from capabilities. Capabilities are the options and opportunities available for individual choice. These characteristics influence personal ability to translate available resources into well-being. A handicapped person may require additional or different capabilities to achieve the same level of functioning comparing to a healthy individual. "The valued functionings may vary from such elementary ones as being adequately nourished and being free from avoidable disease, to very complex activities or personal states, such as being able to take part in the life of the community and having self-respect" (Sen 1997, p. 199). Sen criticized consequentialism for neglecting the value of process itself and ability to select among alternatives, which could contribute to well-being in addition to the final outcomes of actions. Furthermore, individuals may pursue not to benefit from their capabilities, what is precluded in welfaristic model (Hall et al., 2006).

Sen argues against "homo economicus" in favor of individuals with social responsibilities. Personal behavior and choice in neoclassical economics is driven by maximization of individual utility function. Instead, Sen predicates that preferences and actions taken by individuals are driven by notion of potential direct benefit and self-interest as well as humanistic considerations about the societal value of pursued actions (Sen, 1977; Sen, 1985). He believes in "… a variety of human acts and states as important in themselves (not just because they may produce utility, nor just to the extent that they yield utility)" (Sen 1985, p. 22).

Beyond the issues discussed above, welfare framework does not adequately address the questions of social justice. Ethical and equity aspects are assumed to be incorporated into "right" distribution of resources through the political process. In the presence of substantial transactional costs and distortion of incentives by political intervention, efficiency and distributional concerns obviously cannot be separated (Reinhardt, 1992; Hurley, 2000).

Finally, the pragmatic limitation of welfare framework lies in the requirement of a competitive market, which is unattainable for the majority of Public Health activities. Application of Pareto criteria for policy formulation is impractical, as only rare cases

will allow for improvement of some groups without negative impact to the others. Application of a potential Pareto improvement when the gained benefits are substantially large to out weight utility loss, is linked, again, to the not-addressed distributional equity consideration (Hurley, 1998; Hurley, 2000).

Concept of extra-welfarism

Numerous works of Culyer (1990) contributed substantially to development of alternative theory – extra-welfarism. Extra-welfarism builds on Sen's critique of neoclassical framework, specifically, on limitation in assessment of health status using utility function. It is based on two key features: the emphasis on health and rejection of individual preferences to measure social welfare.

1) Emphasis on health.

Extra-welfarism rejects utility and focus on health as the targeted endpoint in theoretical considerations about health care provision. This consideration is essential difference between extra-welfarism and welfarism. Sen, as a well-known critic of welfarism, recommends to focus on functionings and capabilities in normative analyses, rather than utility (Sen, 1987). Drawing on Sen's research, Culyer argues for health as "characteristic" of individuals and suggest to include all non-good and non-utility aspects as sources of well-being (Culyer, 1990; Culyer & Evans, 1996).

2) Rejection of individual preferences to measure social welfare.

Extra-welfarism criticizes individualism arguing that some groups are not always able to assess available options and rationally consume optimal level for their health. Thus, social welfare should not be built on individual preferences alone, what requires involvement of a third party in decision of what to be measured and maximized. This argument conflicts with neoclassical assumption of individualism and consumer sovereignty.

For normative political considerations, extra-welfarism ". . . has taken "health" as the proximate maximand" and suggests to use it as the primer endpoint for health care services (Culyer, 1990). Pareto optimization may be incorporated within extra-welfarism framework, if utility is conceptualizing health measured by quality adjusted life years (Hurley, 2000; Mooney & Russel, 2003).

Critique of extra-welfarism

Critics argue that neoclassical limitations are not adequately addressed in extra-welfarism framework. The main criticism emphasizes the failure to accommodate equity and justice considerations. Efficiency and equity aspects may not be considered separately and introduction of external decision-maker will not imply optimal allocation (Mooney, 2005; Birch & Donaldson, 2003; Hurley, 1998).

The focus on health as primary outcome is another limitation by neglecting other important sources of well-being, like freedom of choice and speech, access to education and others. This failure to accommodate various non-utility and non-health aspects restricts application of extra-welfarism to broader economic activities (Culyer, 1998; Hall et al., 2006).

In extra-welfarism social welfare is determined by a decision-maker, who may over-ride individual preference. Practically, decision-maker may not be capable to decide and achieve optimal allocation by neglecting preferences and choices of individuals or particular groups of people, neither could it be done from justice and equitable point of view (Birch & Donaldson, 2003; Hall et al., 2006). In broader sense, both non-welfarism and extra-welfarism recognize individual and social values, but there is limited operationalization on how these two sets should be combined and considered in maximizing social welfare (Mooney, 2005; Birch & Donaldson, 2003).

Concept of communitarianism

Personal values and preferences play important role in neoclassical and extra-welfarism theories. Many researches oppose to this view pointing out the potential failure of individuals to act rational, to demand and to consume optimally for their health. In extra-welfarism approach, there is also insufficient basis for third party judgment on concept of health, which may vary across different cultures and ethnical groups (Mooney & Russel, 2003). Mooney (1998) suggested to address the above limitations in alternative paradigm – communitarianism - with the emphasis on societal and community role in decisions about equity and allocation of resources in health care. The communitarian framework declines welfarism's assumptions of individualism and consequentialism of the pursued actions and failure to accommodate societal values in individual preferences and utility function (Mooney, 1998; Mooney & Russell, 2003; Hall et al., 2006).

Following aspects are essential premises for the communitarian framework:

1) Non-existence of rationally choosing individuals.

Concept of health and well-being may vary between genders, different cultural and ethnical groups. The use of universal metrics for flourishing and well-being limited to utility function may not be appropriate. The implication of neglecting this aspect may reveal the failure of individuals or groups to desire adequately according to common metrics as well as inability to act rationally in self-interest of their health.

2) Community preferences.

Community plays the central role in the framework and should aid to operationalization of equity and justice concepts within fundamental principles and institutional organization of health care. The weight of equity in the objective function is decided by community and not individuals. Thus, societal preferences and objective function may deviate from totaled individual utility functions.

3) Health as community good.

Communities decide about the needs, allocation of resources and institutions implying health as a societal commodity. Communitarian claims are drawing on considerations of Broome (1991) about individuals taking social role and obligations in the communities they are living in (Broome, 1991; Mooney & Jan, 1997). Here, the equity and justice aspects are getting appropriate weights in the community decision-making about the resource allocation (Hall et al., 2006; Mooney, 2005).

Communitarian view on social justice strongly builds on Sen's arguments about individual capabilities and functionings as central elements in ability to achieve well-being by transforming resources into health and flourishing (Mooney & Russell, 2003; Mooney, 2005; Mooney & Jan, 1997). In communitarianism, individuals can not be seen without the societal context. The objective function contains both individual and community values and preferences. Personal assessment of health may rely on QALY metrics; the appropriate metrics for the community values is yet to be developed. Community is responsible for the provision of health care services to individuals, which are assumed to benefit from the value of the outcome, and consider health care delivery as good itself (Mooney, 1998).

Critique of communitarianism

Communitarian approach ignores non-utility aspects, meaning that non-health characteristics and other process-related features of medical intervention are irrelevant for political decisions. This notion is aligned with QALY assessment in evaluation of social welfare function. However, acknowledgement of health as major objective for social welfare function requires additional analysis to understand the feasibility of the concept in different settings (Mooney, 1998; Mooney, 2005).

Another limitation is related to underweighting or disregard of individual preferences. Communitarianism allow deviation of community preferences from aggregated individual functions. The practical implementation implies involvement of a bureaucratic decision-maker, who will be deciding on objective function, equity considerations and, finally, distribution and allocation of resources. This requires setting of specific rules and assessment of individual-level data about health preferences and status (QALY) with respective weights of competing claims. Collected data should than inform and substantiate preferences and objective function of community. In addition, functioning of communitarian decision bodies would require monitoring of activities by members of community. These requirements are hardly feasible within capabilities of current political structures (Mooney, 2005).

Finally, in addressing equity and justice issues, communitarianism does not provide specific considerations about operationalization for efficiency concept within health care (Hall et al., 2006).

2.4 New Institutional Economics for Studying Political Instruments

In previous sections we stated that (1) Public Health and Economics need common normative theoretical foundation, and (2) Neoclassical economics advocates the need of political intervention to correct market failure, however (3) neoclassical assumption are challenged by non-welfarism, extra-welfarism and communitarianism as not fully valid for normative Public Health research.

Developments in neoclassical economics elaborated different aspects of economic behavior and performance. However, neoclassical view is essentially "institution free", what posses serious limitations for studying health policy within this framework. In addition, ethics, justice, consumer preferences and non-utility outcomes are either implemented in a simplified way or ignored completely in neoclassical framework. Market and public structures are assumed to operate perfectly and costless, firms are "black-boxed" and characterized by production function, all actors have perfect information and decision-makers implement policies "in the public interests". The listed assumptions are not feasible for health care. Thus, normative and positive analysis require out of "black-box" view. New Institutional Economics is filling the vacuum for theoretical analysis of policy, institutions, incentives and efficiency across and within organizations (Robinson, 1997; Hirschman, 1970). Its application to health policy is outlined in following sections.

New Institutional Economics

In majority of the countries health care is regulated and controlled by public institutions. Recent research added substantially to the understanding of the key aspects important for the institutional performance. Principal-agent theory, property rights, transaction cost economics, and public choice theory investigated implications, pros and cons of different forms of institutional arrangements, which are essential for health care analysis. These areas are being developed in a common research stream of New Institutional Economics (Precker & Harding, 2000).

New Institutional Economics seeks to assess and explain the impact and incentives of various socio-economic and political institutions on economic behavior, interactions and performance (Joskow, 2008). Overall, New Institutional Economics suggests that institutions matter and are susceptible to the analysis. New Institutional Economics is a product of evolutionary multidisciplinary work and not an integrated theory, such as neoclassical economics, based on a set of common hypothesis, but rather a combination of different concepts coming from different traditions among which are law, sociology, organizational economics and many others. The foundation was given by works of Ronald Coase (1937b), Douglas North (1990, 2005) and Oliver Williamson (1975, 1985). These individual works, being not fully coherent and consistent, are providing a new perspective and framework for economic analysis (Joskow, 2008).

Principal Agency Theory

Notion of asymmetric information, i.e. different level of knowledge between actors participating in economic transaction, is an essential foundation for the principal agency theory (Ross, 1973; Stiglitz, 1974). Simple principal-agency models investigate maximization of two distinctive utility functions of an ill-informed participant, the principal, and informed participant, the agent. The principal with a limited knowledge and expertise requires agent's involvement and know-how for execution of a particular task, but is restricted or not able to monitor agent's performance and evaluate quality of the final outcomes. The conjoint effect of two specific characteristics – asymmetric information and independent utility functions of principal and agent – is relevant and has major impact in various economic activities and interactions between supervisors and subordinates, insurance companies and clients, patients and doctors (Ryan, 1994; Precker & Harding, 2000). The central theoretical questions in the principal agency theory is the optimal arrangement of incentives and compensation. More generally, specification of explicit and transparent agreements for execution of labor and services on agreed remuneration terms between principal and agent is aiming to align objectives of different utility functions, set incentives and minimize impact of asymmetric information to achieve social and political objectives (Joskow, 2008)

Asymmetric information, distinctive interests and utility functions of interdependent actors are pervasive in health care provision. The relationship between patients, clinical personnel, health insurance, contracting agencies and governmental institutions are classical examples of principal-agent structures. The research of key determinants of utility functions of individual actors, alignment of incentives, rewards and feasible options to monitor the quality of outcomes are central questions to optimize contractual arrangements and so the efficiency of health care provision (Precker & Harding, 2000).

Transactional Costs Economics

The theoretical framework for studying the origination and impact of transactional costs is founded on the works of Coase (1937a, 1937b) and Williamson (1973, 1975) and is strongly related to principal agency theory outlined above. Principal-agent interaction may be influenced by opportunism and bounded rationality behaviors, where transactional costs are the outcomes of deficient contracts (Hart, 1995; 2003). Opportunistic agents will act predominantly in self-interest and pursue the strategies

leading to the achievement of their own goals independently of those of the principal or the organization. Bounded rationality will limit the ability of both principal and agent to make informed decisions and act accordingly due to uncertainty of the actions or outcomes of relevant events and overall complexity of the decision-making process. Comprehensive contracts between principal and agents may describe in details the tasks of involved participants and restrain possibilities for opportunistic behavior. However, taking into account uncertainty of actions and outcomes, the exact formulation of the contractual arrangements may become difficult. Flexible hierarchy and informal settlements may be more practical and efficient, the downside of which is the lack of the detailed specification of the individual roles and potentially more space to act opportunistically. Nonrigid management structure within organization and monitoring activities may help to overcome this tendency. Nevertheless, both specific and flexible contractual arrangements between principal and agents to manage bounded rationality and opportunism will produce some level of transactional costs (Marini & Street, 2007; Precker & Harding, 2000).

The application of transactional costs economics in the health care sector is relevant, since production of health and deliver of health care services is done via complex sequences of transactions among patients, providers, and other stakeholders. Sometimes these exchanges are concrete and observable (e.g. medication is injected), but most of the time the transactions are complex, intangible and abstract (e.g., information and service provision in mixed public-private settings). Information is either incomplete or, conversely, too overwhelming for participants to integrate fully into their decision-making. Numerous public agencies directly regulate or indirectly effect health care delivery and financial transactions. High entry barriers in health care markets make it prohibitively costly for new organizational arrangements. Actors pursue their own interests, often professionally justified, at the expense of their exchange partners. Mismatches occur between objectives and allocated resources. Finally, the environment is unpredictable and highly uncertain (Stiles & So, 2003). Application of transactional costs economics in health care often refers to analysis of the optimal contractual arrangements within public and private organizations in health care context (Bartlett, 1991; Ashton, 1998; Keen & Ferguson, 1996; Hughes et al., 1997).

Property Rights Theory

Property rights theory investigates effect of private and public ownership on incentives, efficiency and broader economic activities of the public and privately own organizations. This theory has mutual conceptual roots with principal agency and transaction costs theory (Kim & Mahoney, 2005).

Private ownership has historically proved to have strong positive incentives on efficiency. Property rights theory investigates this phenomena from the perspective of residual decision rights and the ability to allocate the residual returns. The possession of residual decision rights allows to determine the allocation of assets or resources, and the ownership of residual return claims entitles to decision about the distribution of profit or loss (Milgrom & Roberts, 1992; Precker & Harding, 2000).

There are different forms of property and ownership. The owner of an asset may or may not be in possession of both residual decision rights and claims. The implications of possession of residual rights may be obvious for the commodities, such as a vehicle or house, but become much more complicated if the effects are considered within multilayered organizations. Large companies and public institution may accommodate or be entitled to administration of big assets. The relationship between owners of assets and those having rights for the decision about residual revenue may have complex effect on incentives and efficiency within organizations and sharpen principal-agent dilemma (Precker & Harding, 2000).

Property rights theory has strong application in health care sector to study incentives and efficiency of the whole system and particular branches. State's subsidization and various forms of interference may induce inefficient operation and production within non-profit organizations comparing to commercial competitors. Empirical foundation of this proposition can be found in analysis of different data sources from US health care market, e.g. HMOs (Frech, 1976), assisted living facilities (Frech, 1985; Tuckman & Chang, 1988), public and private wards and clinics (Clarkson, 1972; Friedman & Pauly, 1981; Bruning & Register, 1989; Register et al., 1991).

Public Choice Theory

Theoretical considerations about public choice in political and governmental context were developed by James Buchanan (1963), George Stigler (1971) and Sam Peltzman (1976). The research of George Stigler, James Buchanan, and Gary Becker were awarded the Nobel Prize in Economics.

Public Choice theory investigates performance, efficiency and optimal size of governmental structures and policy. This field of research elaborates the impact of the self-interested behavior of politicians, interest groups, and bureaucrats on public policy. Bureaucrats are seen as maximizers of their own utility function, which can be expressed in term of assigned administrative budget. The theory suggests, that under particular conditions, the politicians may attempt to get administrative rights over higher share of national resources than is needed to provide specific service or execute their central functions. In fact, government or administrative structures may "outgrow" the needed size, bind or inefficiently allocate available resources. As a result, governmental bodies will potentially fail to perform their core-functions and introduce additional distortion into economic activities (Precker & Harding, 2000).

Public choice theory uses analytic tools such as decision and game theory to uncover and explain causes and trends for inefficient bureaucratic structures and decision-making. The failure of political interventions is rooted in existence of asymmetric information, bounded rationality, opportunism and transactional costs. Public choice theory is often invoked in discussions of limiting state's intervention and liberalization of market activities and transactions (McPake & Normand, 2002).

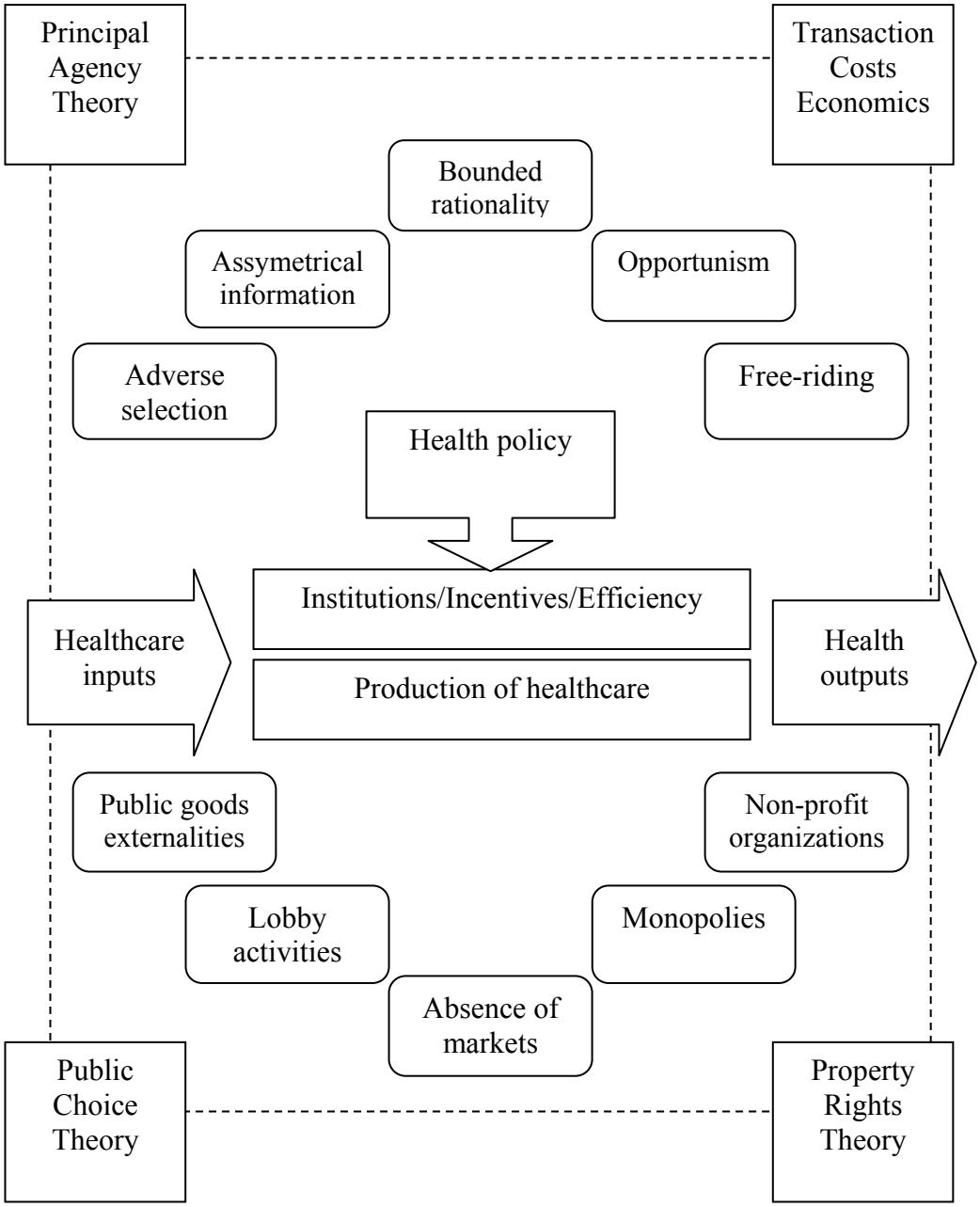
Public choice theory finds broad application in health care sector, where drug approval process is heavily regulated by such bodies as FDA and other National Health Authorities. In UK reimbursement decisions and coverage issues of drugs and therapies are subject of influential activities by different actors, such as NHS, NICE, PCTs, patient groups, industry, academics, medical experts etc.

Health Policy is steered and influenced by multiple institutions

Neoclassical theory justifies political interventions in health sector to secure allocative efficiency, which is achieved when resources are allocated to maximize the welfare of the community (Drummond, 1991). Health care "production" process, as depicted in

Figure 3, is essentially black-boxed in neoclassical framework. Branches of New Institutional Economics, including public choice, transactional costs, principal agency and property right theories, provide a theoretical basis to study essential aspects of health market such as asymmetric information, free-riding, monopolies, adverse selection and others. Specifically, New Institutional Economics explores the incentives of regulations and institutions, which influence efficiency of the transforming resources in terms of capital and labor inputs into health care outcomes. With regard to generalizability issue, new institutional economics suggests that institutions may moderate or have direct impact on health economic outcomes. Geographic variability in institutional arrangements of health care provision and regulations may produce different health outcomes at different costs for the same intervention for comparably homogeneous clinical population.

Figure 3: Theoretical Aspects of New Institutional Economic for Health Policy



Source: compiled by the author

2.5 HTA and Health Economic Evaluations for Informed Health Policy

Previous sections reviewed main economic theories relevant for Public Health and policy research. We stated that (1) from normative view, Public Health and Economics should not be studied explicitly apart from each other, (2) neoclassical economics advocates the need of political intervention to correct market failure, however (3) neoclassical assumption are challenged by non-welfarism, extra-welfarism and communitarianism as not fully valid in Public Health context, and, finally, (4) New Institutional Economics suggested the major role of institutional arrangements for studying approach, implementation and efficiency of political measures. Our analysis failed to identify common theoretical foundation between Public Health and Economics, however, there seems to be consensus about the need of public intervention to promote health in the community through thoroughly designed health policy. The instrumental concept to inform, implement and monitor health political decisions is Health Technology Assessment. This concept is rooted in multiple Economic and Public Health theories. Generalizability methods are analytic tools to produce evidence for HTA. The concept and requirements of HTA are summarized below.

Health Technology Assessment

Historically, the outcomes of health interventions were often determined in uncontrolled experiments, individual cases and by opinions. Admittedly, availability of several treatment alternatives for the same clinical indication is fairly recent phenomenon. Few decades ago there were only selected technologies feasible for comparative assessment (Jonsson, 2002). After technological progress advanced medical care and provided clinicians with new treatment options, health care systems in industrialized countries faced rising expenditures in perceiving the dissemination of new technologies to provide high-quality services to broader population. Even today, almost any health innovation is welcomed as a potential treatment for a disease. Several new treatments, however, failed to provide substantial benefit (Jonsson & Banta, 1999). The question of costs, effectiveness and quality of the clinical interventions and measures come up.

Health technology assessment measures the impact of the treatment or diagnostics procedures on society and provides scientific basis for acceptance, modification or rejection of technologies. The central research question addressed by HTA investigates whether a new technology is superior to existing alternatives in terms of clinical, economic and social outcomes (Hutton et al., 2006). The concept of HTA is sometimes misleadingly associated with the rigid cost-control mechanisms, which is not its primary purpose. HTA is aimed to contribute to efficient use of limited resources under considerations of quality, safety and equity. The improved efficiency within the health care system will enable to stimulate, reward and extend the access to the valuable innovations (Jonsson & Banta, 1999; Kulp & Greiner, 2006).

Acceptance and implementation of HTA is difficult due to different values and opinions of the involved parties - decision makers, clinicians, patients, providers, payers and other actors. This can be partly due to the fact that the conclusions of HTA do not always confirm clinical opinions or even disenchant expectations. It is also not universally accepted, that assessment of health technologies should be done from perspectives other than that of clinicians (Jonsson, 2002). HTAs are designed to inform decision makers, who are assumed to act in the interest of general population. The opposite view have many key-players, in particularly clinicians, who are educated and trained in spirit of caring for and addressing the needs of individual patients (Rosen & Gabbay, 1999).

Implementation of HTA in different countries

In 1992 Australian Pharmaceutical Benefits Advisory Committee was the first regulatory body which implemented HTA in the decision-making process regarding reimbursement of a new medical technology. Since then, many European authorities elaborated and established application of country-specific health economic guidelines and standards. The official formation of HTA decision-making body was initiated in UK in 1999 with establishment of National Institute for Clinical Excellence (NICE).

Following UK example, several European countries institutionalized HTA bodies in political decision-making process. Submission of health economic evidence is now mandatory in Belgium, Scotland, Netherlands, Norway, Spain, Portugal, Poland, Hungary and other countries. The extent, to which HTA influences policy and resource distribution in health care, differs between countries. Variation in decision-making

process and administrative bodies can be partially attributed to differences in underlying political and institutional organization. HTA authorities in Scotland, Belgium and Sweden, like NICE, go beyond clinical efficacy and demand evaluation of health economic impact from broad societal perspective. In Italy and France the evaluation of unmet need and budget impact analysis is done by different regulatory bodies (Hutton et al., 2006). In Germany, HTA evaluation is governed by Institute for Health Care Quality (IQWiG). German methodological requirements, however, depart from the international requirements and are being criticized by academic research. In general, with all variety of forms and structures, HTA and health economic evaluations are now formally or informally used to inform health policy decision-making.

Methods for health economic analysis

The most relevant types of health economic analysis are the following :

- cost-of-illness analysis
- cost-effectiveness analysis
- cost-utility analysis
- cost-benefit analysis.

Cost-of-illness analysis determines overall socio-economic burden of clinical condition in estimating associated direct, indirect and intangible costs.

Cost-effectiveness analysis is a comparative analysis of feasible treatment alternatives. In this type of analysis the difference in costs between new and standard therapy is related to differences in effectiveness. The ratio may differ in the magnitude between alternative programmes. The result is stated in the terms of costs per additional unit of effectiveness , e.g. costs per additionally cured patient.

Cost-utility analysis in it structure and approach is analogous to cost-effectiveness analysis. Here, the incremental costs of a program are compared to incremental health improvement, where the health improvement is measured by quality-adjusted life years (QALY). The outcomes are expressed as costs per QALY.

Cost-benefit analysis values resources and benefits of different medical technologies or interventions. The result of such analyses might be stated either in the form of a ratio

of costs to the benefits, or as an absolute positive or negative value depicting the net loss or benefit of new intervention comparing to alternative (Goodman, 2004; Drummond et al., 2005).

Theoretical foundations of health economic methods

The above methods are now the standard tools within health economic evaluations. However, the theoretical foundations of these methods are not fully established. Cost-benefit analysis compares the net costs with the net benefit. Taking welfaristic point of view, only individuals can appropriately valuate their own health status and the utility of the new treatment. Thus, individual willigness-to-pay is legitimate measure of the benefit including those of health care. Cost-benefit analysis justifies the implementation of the interventions which raises the net benefit, what is clearly welfaristic approach to the analysis (Dolan & Edlin, 2002; Johannesson & Karlsson, 1997).

Contrary to cost-benefit analysis, the theoretical foundation of cost-effectiveness and cost-utility analysis is less clear. These forms of analysis are often criticized by economists within the health sector because of the difficulties (conceptual, ethical, practical) in monetarizing life-years gained, as well as by extra-welfarists who emphasize the health as primary outcome for normative analysis in the health sector. Cost-effectiveness analysis compares the incremental costs with incremental effectiveness, i.e. how change in costs is counterbalanced with gain or loss in clinical effectiveness. The costs are measured in monetary terms and the effectiveness in the natural units. The outcomes are presented in term of cost-effectiveness ratio, i.e. additional costs per additional unit gain. In cost-utility analysis, costs are again measured in monetary units, but the outcome is measured in terms of quality-adjusted life-years (QALYs) that reflect both the quantity and quality of life years gained as a result of the intervention (Hurley, 2000). Typically, the programs with cost-per unit outcome below particular threshold – shadow price or willingness-to pay of the decision maker - will be funded. For a cost-utility analysis that uses utility-based QALYs as the measure of outcome, the theoretical considerations could be linked to communitarianism (Hall et al., 2006). The normative foundation of cost-effectiveness analysis with measures incorporating non-utility weights may be found in extra-welfarism (Garber & Phelps 1997; Hurley, 2000).

Requirements for informed decision-making

The formal implementation of health economic analysis to inform decisions about the reimbursement of new treatments or technologies in different countries possess substantial challenge on health economic data and methods. Sculpher and colleagues (2006b) and Claxton and colleagues (2002) outlined following major requirements:

1) Precise definition of research purpose.

Clinical patient population must be clearly defined in any economic analysis and all available and relevant treatment options should be identified. The specification of relevant treatment options may involve comparison to both on- and off-label technologies and interventions, as well as different sequences, combinations and dosing of therapeutic and diagnostic procedures.

2) Explicit specification of the measurement of health benefits.

Many coverage decisions require maximization of a specific objective function constrained by available budget within the health care system. The objective function is likely to optimize particular health outcome, measured either in natural units or in quality adjusted life years. Concept of QALY is broadly used in economic analysis of various technologies and indications. However, researchers need to consider and include all outcome measures relevant to the patients and clinical practice. The assessment of health gain should be clearly explained in order to evaluate the "level of substitution" comparing to health outcomes under alternative treatments. This is essential to maintain consistency in analysis and decision-making in rejecting or adopting new technology.

3) Sufficient timeframe.

Important requirement in health economic assessment is the need to evaluate the effects and implications of the new technology over the relevant time period within major costs and clinical outcomes are likely to occur. Sufficient time horizon may often extend over a life-time period, particularly when clinical events are rare, condition is chronic and or have effect on mortality.

4) Inclusion of comprehensive evidence.

Analysis based on incomplete evidence may come to ambiguous outcomes and misleading recommendations. Treatment effect should be compared to the full range of

the treatment options used both as indicated and off-label, if feasible. Valuation of resource use and costs should avoid assessment of selected episodes or resources, as it may ignore potential costs "shift" between the services (e.g. from stationary to ambulatory care).

5) Relevance to the jurisdiction of interest.

Health economic decisions are essentially jurisdiction-specific and so is the need of health economic data. The central generalizability question remains important in assessment of the applicability of external evidence to the context of a specific jurisdiction.

6) Evaluation of uncertainty.

Uncertainty is related to the imprecision in estimation of the relevant clinical and economic outcomes. The correct quantification of uncertainty in parameter estimates is essential to evaluate the need for additional research and the probability of making the "wrong" decision based on current evidence. The implications of making "wrong" decision will result in inefficient resource allocation and lost opportunity to health gain using alternative treatment options (Sculpher et al., 2006b; Claxton et al., 2002).

Limitations of evidence available for decision-making

Decisions about reimbursement of a new technology are often made based on incomplete evidence. The main vehicles for comparative health economic analysis of a new technology are the patient-level health economic studies conducted along phase III or phase IV clinical studies, and decision models, which use data on treatment efficacy and resource use from the patient-level trials. Thus, health economic analysis is inevitably influenced by clinical development, what posses several methodological challenges.

First challenge is the focus of clinical trials on maximization of internal validity, i.e. the estimation of treatment effect as unbiased reflection of the observed value in the trial participants in a strictly controlled environment (Mason & Mason, 2006; Kinosian & Glick, 1998). The hierarchy of clinical evidence suggest the strongest internal validity in well-designed, randomized, blinded clinical trials followed by nonrandomized trials, prospective/retrospective cohort studies, prospective/retrospective case-control studies and, finally, opinions and beliefs, which should be considered as of very limited internal

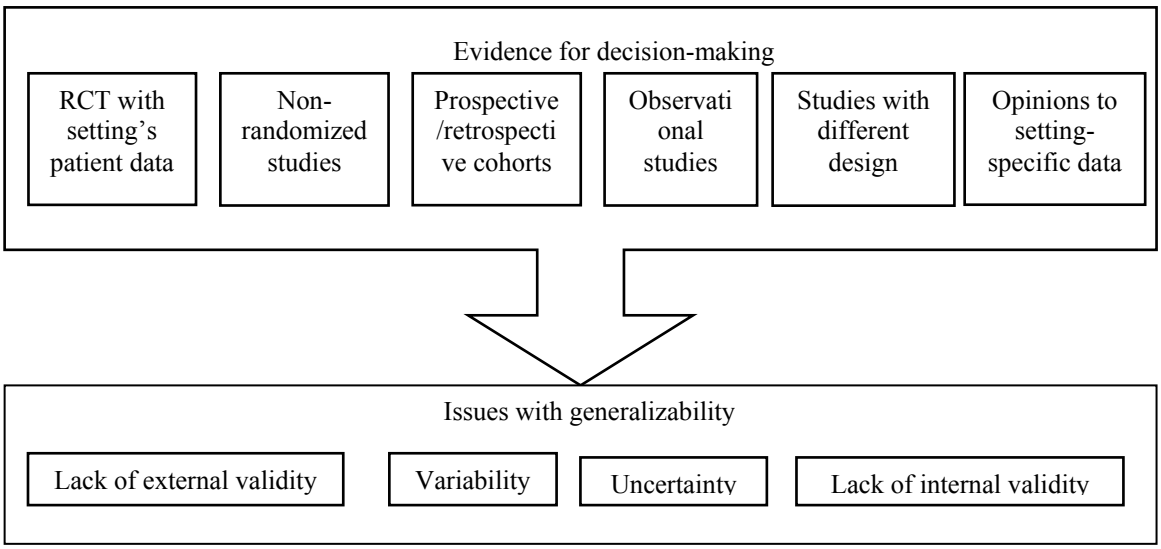
validity (Goodman, 2004). Health economic analysis deems to estimate resource utilization in a real-life setting with emphasis on external validity. Thus, there is a certain trade-off between internal and external validity and health economic analysis should balance both (Jonsson & Weinstein, 1997).

Second challenge relates to uncertainty and variability of the clinical and economic estimates. Economic literature is not aligned in defining these concepts. In following we apply the taxonomy described by Coyle (2003) and Thompson (2002). Variability reflects the differences in the parameters across the population or sample. It relates to the randomness of the event in the population and cannot be reduced. Rather, more information can allow stratification into more homogeneous population groups. Uncertainty expresses the imprecision or lack of information regarding the true value of the outcome in the given population. Variability and uncertainty are distinctive. Uncertainty can be addressed through better measurement e.g. additional studies; although, further knowledge may fail to reduce uncertainty over the optimal decision. Knowledge about uncertainty, unlike variability, can be represented by probability distributions (Coyle, 2003; Thompson, 2002).

Improving generalizability of available evidence for informed decision-making

Generalizability of health economic evaluations for informing decision in particular context is dealing with both internal/external validity and variability/uncertainty issues (see Figure 4). The evidence available to the decision-making can be collected in randomized clinical studies, not randomized studies, observational cohort studies and finally by individual clinical judgement. Different research designs and methodology have strengths and weakness. E.g. randomized clinical studies are considered as “golden standard” for assessment of clinical efficacy, but may lack of external validity due to restrictive selection criteria and procedures irrelevant or not feasible for routine clinical practice (Bombardier & Maetzel, 1999). Observational studies may overcome limitations imposed by strict clinical protocol for randomized clinical studies, but may fail to control for possible confounding or selection bias resulting in high uncertainty around the clinical and economic outcomes. Finally, individual clinical judgements and opinions may have limited validity, if based on few selected observations.

Figure 4: Evidence Required for Decision-Making in Health Policy



Source: compiled by the author

In broader sense, greater understanding of generalizability of economic evaluations has potential utility for both health research and policy. Three main groups could stand for benefit: researchers, decision-makers and interdisciplinary community involved in health care provision (clinicians, pharmacists, psychologists etc.).

Relevance of generalizability methods for Public Health, policy and research

Decision makers are concerned with the service provision in the settings of their responsibility. In order to decide about reimbursement of new intervention, local authorities often face important question: do the results of multinational health economic study apply to the local context? Given the number of various levels in different health care systems, it will be seldom possible to conduct a customized study for a given setting and policy question. Generalization and interpretation of the results of available studies is inevitable (Sculpher et al., 2004).

HTA contributes to dissemination of politically supported technologies (Kulp & Greiner, 2006). However, it is also recognized, that HTA and health economic analysis still have little impact on the political decision making in health care, leaving a gap between what we know and what we do (Rosen & Gabbay, 1999). This problem could be partially explained by the fact that health economic studies may not be conducted for all possible jurisdictions. Additionally, there is a time lag between introduction of a new

technology and availability of the economic evidence from use of this technology in the clinical praxis. Therefore, the limited economic data from international studies should be utilized using generalizability methods to inform decision-makers at the early stage of the technology introduction.

As mentioned earlier, broader application of economic evaluations (incl. generalizability methods) will contribute to promotion and implementation of public health programmes. Assessment of the outcomes of public health interventions is important, as in absence of economic data, decision makers might favour programmes with good evidence on cost-effectiveness from randomized clinical trial for e.g. medical treatment and abandon the programmes, where no economic evidence is available (Rychlik et al., 2005; Banta, 2003).

Researchers will benefit from the development of generalizability methods, which will extend the understanding about causes of variability in costs and outcomes between settings and it implications for other settings. Potentially, generalizability methods will increase utilization of data, methods and results of the available studies. This knowledge will help to avoid redundant studies and save research fundings (Sculpher et al., 2004). In addition, efficient utilization of available data will help to formulate future research to close critical knowledge gaps. Furthermore, economic perspective and outcomes of many actions in the area of New Public Health might be difficult to evaluate such as those in health promotion or environmental improvement (Rychlik et al., 2005). Development of generalizability methods will increase the methodological stock, which should be used to assess and quantify economic outcomes for public health programmes.

Essentially, generalizability methods should help to provide jurisdiction-specific evidence to inform local policy, support promotion of HTA in policy formulation and implementation, increase evidence stock, efficiency of resource allocation and formulation of future research.

2.6 Summary of Chapter 2

This section embeds and links generalizability methods for international health economic studies to the theoretical and instrumental concepts in Public Health and Economics. We reviewed theoretical paradigms mainly from economic perspective starting from neoclassical mainstream economics followed by newer theories such as non-welfarism and New Institutional Economics, linking these to the instruments of health policy. Summary of the conclusions is presented below.

Positive and normative analysis is a foundation for Public Health instrumental concepts such as health economic evaluations and generalizability methods.

Positive economic analysis describes the effect of the political intervention on a particular setting. Normative economic analysis attempts to outline the "ideal" policy with desired socio-economic outcomes. Additionally, normative analysis should guide the development of evaluation methods to assess and interpret effects and outcomes of various health care interventions. The major underlying theories of normative health economics are neo-classical welfare theory (welfarism), extra-welfarism, communitarianism and New Institutional Economics.

Health policy is anchored in Economic and Public Health theories. Neoclassical economics advocates political intervention to improve allocative efficiency.

The mainstream economic theory agrees that the "invisible hand" of the competitive market will not necessarily allocate limited resources to get maximum health gain, due to violation of free market conditions, attributed to asymmetric information, uncertainties and externalities. Thus, imperfection of free-market in health care creates an important role for non-market institutional arrangements (i.e., the political visible hand) to compensate for free market failure. In order to improve health care system performance (and not to do more harm than good, as competitive market incentives are applicable for many health care sectors), political decision-making should be informed by setting-specific evidence.

Neoclassical Economics is not fully aligned with Public Health perspective on ethics, health, capabilities and well-being. Non-welfarism, extra-welfarism and communitarianism are emerging theories toward Public Health concept.

The framework of welfare theory is based on the utility maximization, consumer sovereignty, consequentialism and welfarism. Social welfare is an aggregated function over individual utilities, whereas the source of individual utility is generally restricted to consumption of goods and services. Welfarism focuses on efficiency and neglects ethics and equity issues. These assumptions are broadly criticized by researchers. Alternative non-welfaristic framework, developed by Sen, is particularly influential in health economics. He emphasizes importance of individual characteristics and capabilities as determinants of personal well-being along with considering social responsibilities in individual objective functions. Sen advocates duality of health as a capability and a characteristic. Alternative theories, such as extra-welfarism and communitarianism, build on Sen's work. The considerations of justice should target both access to health care resources and distribution of health.

New Institutional Economics provides a complementary theoretical basis to neoclassical economics to study efficiency of health production, incentives and impact of institutions and policy.

Health care provision is strongly regulated and influenced by public institutions (organizations and political instruments), whereas neoclassical model is essentially institution-free. New Institutional Economics is filling the theoretical gaps and provides complementary framework to study incentives and effects of health policy and institutions in health care. This can potentially expand our understanding about geographical variation and efficiency of different health care systems and, thus, improve generalizability of health economic evaluations conducted in different health care settings.

Theoretical foundation for health economic methods in Public Health and policy are yet to be elaborated.

Cost-benefit analysis is rooted in neoclassical economics; the theoretical foundation of cost-effectiveness and cost-utility analysis is less clear. There is no universal agreement between economic theories such as non-welfarism, extra-welfarism, communitarianism

and others with Public Health paradigm on essential questions to define health policy. For instance, there are different views on rationality and preferences in the decision-making, whether and how individual and population health and welfare should be defined and measured, what ethical, equity and social justice considerations should be targeted or taken into account while formulating health policy. These and other questions merit further elaboration of a common theoretical foundation, which clearly require a multidisciplinary approach of Public Health researches and unlikely to be developed individually in "established" biomedical and social sciences.

Generalizability methods are relevant for Public Health research and policy. These methods will (1) contribute to better understanding of efficient health care provision and (2) increase evidence body for jurisdiction-specific health policy.

Assessment of geographical variability contributes to understanding of efficient health care provision and aid to formulation of political interventions. Apart from differences in economic development across geographic areas, different understandings and interpretations of the health policy objectives lead to various institutions, political decisions, resource allocations and even empirical indicators measured within different systems. Improved generalizability of empirical data and comparison of international health economic evidence will help to evaluate efficiency of health care provision in different settings.

Studying variability and improving generalizability of health economic evidence is essential to inform jurisdiction-specific health policy, which, in turn, should improve health and reduce inequalities. HTA is not yet fully established in all settings. Conducting of health economic evaluations is usually subject to budgetary and time constrains. Consequently, available evidence to inform health policy measures is often scarce and not always collected in the jurisdiction of interest. Improving of generalizability of available health economic evidence for jurisdiction-specific decision-making will help to save resources dedicated to health care research and formulate future research needs.

3 Methods to Improve Generalizability of Multinational Health Economic Evaluations

In the previous section we outlined theoretical foundations related to Public Health, policy and HTA, justified political intervention in health care market and emphasized importance of institutional arrangements. Immatureness of normative view and theoretical paradigm in Public Health contributes to differences in political agendas, instruments and institutions across geographic health care settings. The institutions play essential role by setting regulatory environment and influencing incentives and efficiency of health care production. This, in turn, may enhance variability in input and output parameters between different health care systems, which may not be fully explained by differences in absolute level of economic resources available in the particular setting. The implication for generalizability of health economic evaluation is that the outcome of a specific treatment in economic and even clinical measures may not be considered explicitly apart from the health care setting where it has been produced.

The purpose of this section is to summarize current approaches to address heterogeneity and improve generalizability. We conducted targeted review of health economic literature and international guidelines. Based on this review, we selected and adjusted analytical methods for the case study. This chapter is structured as following. We first discuss terminology and describe search strategy applied in literature review. We then summarize evidence related to variability factors, requirements and recommendations from international health economic guidelines, analytical approaches used in decision models and patient-level studies. Finally, we describe generalizability methods applied in the patient-level case study.

3.1 Terminology and Literature Search Strategy

Understanding of the concept and terminology, used in the literature in relation to generalizability issues and methods, is required to specify targeted literature review strategy. These topics are summarized in the following.

Terminology

There is no universal agreement about generalizability definition in the literature. Generalizability has been interchangeably used with other terms and concepts, such as transferability, portability, external validity, extrapolation and others (Mason & Mason, 2006).

The term "transferability" is generally used to compare results of health economic evaluations across geographic locations (Wilke, 2003; Greiner et al., 2000). Geographic "transferability" is of particular interest for health policy in the cases, where cost-effectiveness data for the specific settings or country is not available. The term "extrapolation" is often attributed to the time aspect of the evaluation in e.g. modelling of long-term treatment effects. "External validity" is discussed in conjunction with "internal validity" of the outcomes generated in randomized clinical trials, where treatment efficacy is proved with specific procedures and methods on a selected patient population (Evans & Crawford, 2000). "External validity" may be seen in epidemiological context as the extent, to which the results of clinical trials are applicable to and valid for the everyday medical practice (Baltussen et al., 1999; Coyle & Lee, 1998).

For the purpose of this thesis, we use broad concept of generalizability suggested by Sculpher and colleagues (2004). They define generalizability as "…the extent to which the results of a study, as they apply to a particular patient population and/or a specific context, hold true for another population and/or different context "(Sculpher et al., 2004 p. X). The scope of this definition includes geographical, time and epidemiological components.

Objectives of literature review

We conducted targeted review of the health economic literature and international guidelines published between January,1997 – January, 2009. Specifically, review had following objectives:

- To **specify variability factors**, which affect generalisability of the results of economic evaluations across geographic areas and over the time.
- To identify **approaches** to address generalisability suggested by the international health economic and methodological **guidelines**.
- To review **essential methods** to improve the generalisability of evaluations based on **decision models**.
- To review **essential methods** to improve the generalisability **of patient-level studies.**

Search strategy

Collection of the evidence relevant to generalizability aspects was done in targeted way as following:

- Literature published in January, 2004 – January, 2009 was searched in PubMed, Google, various health economic sites, ISPOR website and health economic conferences.
- PubMed search was conducted on 4th of March 2007 and 2nd of February 2009 using following query terms: (("economics"[TIAB] NOT Medline[SB]) OR "economics"[MeSH Terms] OR economic[Text Word]) AND (generalizability[All Fields] OR transferability[All Fields] OR generalisability[All Fields]) AND English[lang] AND medline[sb] AND "humans"[MeSH Terms]).
- Literature published between January, 1997 – January, 2004 was retrieved from the study of generalizability issue conducted by Sculpher and colleagues from York University, commissioned by National Health Service HTA Programme in UK (Sculpher et al., 2004).

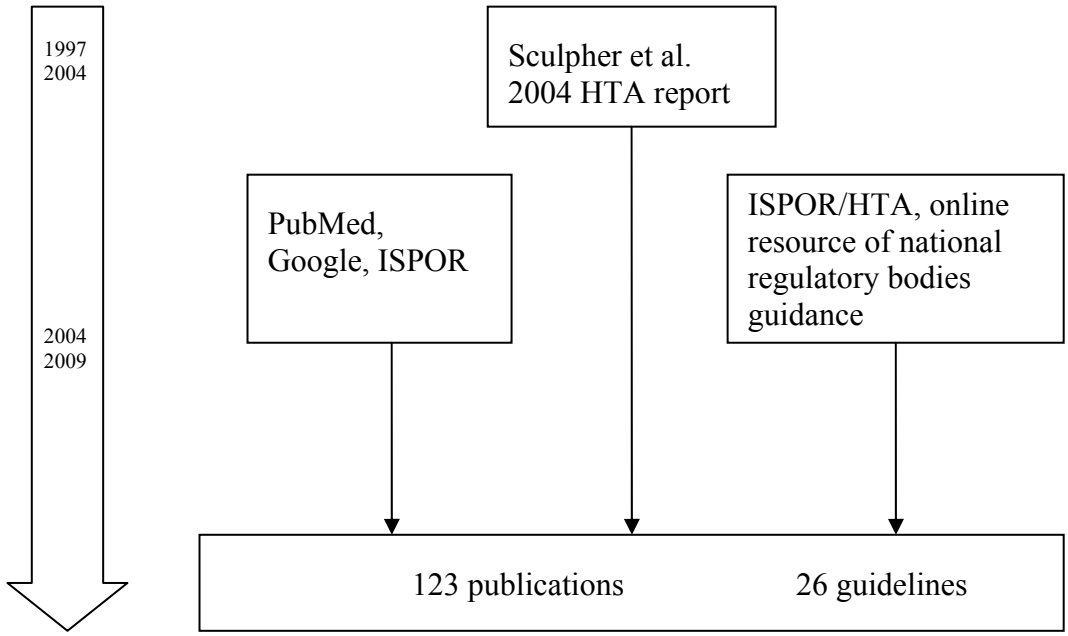
Publications were included based on the following inclusion criteria:

- Health economic publications in English language.
- Health economic/methodological/regulatory guidelines.
- Methodological publications on generalizability or transferability methods.
- Multinational patient-level studies and health economic evaluations based on decision models, which explicitly discussed generalizability and/or transferability issues.

Included literature

123 publications and 26 international health economic guidelines were included into review. Literature search strategy and included evidence is graphically presented in Figure 5.

Figure 5: Literature Search Strategy



Source: compiled by the author

Limitations of the search strategy

This review has several limitations. It can not be considered as complete and exhaustive due to the fact, that only recent published health economic evaluations in English language were retrieved. Only the publications, which explicitly discuss

generalizability or transferability issues were selected and reviewed. All multinational studies, which potentially could have applied generalizability techniques without referring to generalizability term, may not have been identified and included into the review. Additionally, potential generalizability methods outside health economic analysis in e.g. epidemiology, education, social and political sciences were out of the scope of this literature review.

3.2 Factors Contributing to Between-Country Variability

The need to develop generalizability methods evolved in discussion of observed variability in cost-effectiveness outcomes across locations and over the time. A fairly extensive literature exists on the range of different factors that could influence variability in cost-effectiveness results of different treatments. The detailed overview of variability factors is provided by Sculpher and colleagues (Sculpher et al., 2004, chapter 3). Our literature review of the recent publications published in 2004 – 2009 confirmed these findings. The most frequently discussed factors in the health economic literature can be divided into patient, disease, health-care system and methodological factors as summarized in Table 1 below.

Table 1: Variability Factors within Health Economic Evaluations

Group	Factors	Description
Patients factors	Demographic and clinical characteristics	Demographic (age, gender, race), co-morbidity, risk factors, genetics, medical history, life expectancy.
	Socio-economic characteristics	Income, lifestyle, educational level, religion, employment, migration, traveling, living conditions, nutrition, individual preferences for particular levels of health (reflected in utility values of health status), acceptance of technology.

Group	Factors	Description
Disease factors	Clinical presentation	Epidemiology (incidence/prevalence, spread), disease severity and progression, mortality, interaction with other conditions and co-medications.
	Clinical practice	Clinical conventions, diagnostics and treatment guidelines, applied treatments.
Health-care system factors	Health care resources	Level of economic and technological development, social security, access to resources, environment, absolute and relative prices, regulatory system, organization of delivery system e.g. staff, facility, equipment, access to advanced technology/treatment, capacity utilization.
	Remuneration and incentives	Types of coverage public/private/mixed, user fees, co-payments, deductibles, incentives for patients, providers and suppliers, methods of remuneration, availability of generics and substitutes.
	Provider characteristics	Type of facility, specialization of labor, quality of care, technical efficiency, capacity utilization, level of competition.
Methodological factors	Study design and conduct analysis	Protocol-driven procedures, industry bias, artificial study conditions: placebo-controlled, improved treatment compliance, not-representative settings, selection of sites, treatment comparator, estimation of clinical and economic outcomes, costing methods.
	Timing	Timing of economic evaluations, learning effects, technological innovation, advancement of technology/practice, development and availability of new diagnostics and adjuvant treatment options.

Source: modified from Sculpher and colleagues (2004) and Goeree and colleagues (2007).

3.2.1 Patients' Factors

Major source of variation in cost-effectiveness is attributed to the patients' characteristics, which could vary within and across jurisdictions and can effect the ability to generalize cost-effectiveness results across geographic areas. Patients' case-mix is comprised of clinical and wider socio-economic characteristics.

Clinical patient characteristics

These characteristics include age, gender, race, risk factors, co-morbidities, genetic factors, life expectancy. For instance, age can have an impact on cost-effectiveness, as it is directly associated with susceptibility to severe conditions, such as diabetes and myocardial infarction. Severe chronic co-morbidity (e.g. HIV) can affect clinical outcomes of the treatment. Combination of clinical factors can affect baseline risk of disease and ultimately impact clinical outcomes and economic resource use (Goeree et al., 2007). Life expectancy has a direct impact on QALYs gained, especially for the treatment of chronic conditions (Welte et al., 2004). Characteristics of the patients vary strongly across locations and the difference on this level may affect presentation and course of condition, and thus, have effect on outcomes and resource use. Stratification of patients into relevant sub-groups is important in health economic evaluations in order to inform political decisions about the groups of patients who will mostly benefit from a new treatment (NICE, 2008; Raisch et al., 2003; Bryan & Brown, 1998; Ghatnekar et al., 2001).

Socio-economic characteristics

Number of non-clinical factors are hypothesized to have effect on resource use and costs and vary across locations. In broader terms, income and education could influence access to the preventive, diagnostic and treatment options (Tsuchiya et al., 2002; Welte et al., 2004; Raisch et al., 2003; Sculpher et al., 2004). Treatment effectiveness for several chronic conditions and thus, the results of cost-effectiveness analyses will depend critically on patients lifestyle, attitudes, and behavior (e.g. diet, sport exercising, smoking and alcohol consumption). Acceptability of new intervention has impact on compliance and final outcomes. The level of acceptance may vary with cultural and religious preferences for e.g. use of contraception, screening, vaccination and organ

transplantation. In addition, compliance may be affected by educational levels and cultural factors, which vary across the setting. The implications of non-compliance are likely to be reflected in increased resource use and or reduced effectiveness. Finally, compliance or adherence to procedures achieved in clinical trial may not be replicated in routine practice. Health-related preferences used to calculate QALY in cost-utility analysis is another source of between-country variability. Different utility weights used for different countries (e.g. EQ5D utility score) will produce different study outcomes and may prevent from generalizability to external settings. The underlying factors inducing this variation have not yet been extensively studied and need further exploration (O'Brien, 1997; Sculpher et al., 2004).

3.2.2 Disease Factors

The second broad category of factors related geographical variability is comprised of clinical presentation of the condition and clinical practice of the disease management.

Clinical presentation

There is number of geographical characteristics, that can affect the incidence and prevalence of the disease, such as presence of specific genetic factors (e.g. for sickle cell anemia), environment (UV light and skin cancer), migration and traveling (hepatitis, HIV, tropical diseases) (Welte et al., 2004). Interventions focused on prevention are more likely to be cost-effective with higher incidence of a "target" disease (Drummond et al., 1997).Variability in patients' characteristics, cultural and environmental differences may affect disease severity or case mix in one country compared to another. Similarly, disease progression and spread in the case of infections may vary across jurisdictions depending on prevention programmes, environmental factors, population density, life-style, treatment acceptance, and notification programs. Finally, disease-specific mortality can vary across geographic regions (Reed et al., 2005; Goodacre & Dixon, 2005).

Clinical Practice

Routine medical practice, diagnostics, treatment guidelines and conventions are known to be differently implemented within and across the countries. In part, these differences might be attributed to the specific organization of health care provision

according to local providers' structure and characteristics of service delivery. For instance, in France specialized medicine is generally provided by community-based clinicians, whereas in Great Britain the major part of the services is institutionalized within hospitals. Different diagnostic approaches and treatment patterns have effect on demand, resource use and relative costs of procedures. This, in turn, will affect supply of the health care services and access to new diagnostics and treatments. In addition, differences in available treatment alternatives will affect local clinical practice and conventions. At a very minimum, variability in clinical practice is reflected in different option for comparative and standard treatment, what may produce different cost-effectiveness outcomes (Sculpher et al., 2004; Mason & Drummond, 1997; Revicki & Frank, 1999).

3.2.3 Health Care System Factors

Price weights are known to differ between and within countries. Most researchers advocate at least replacement of a unit prices, when transferring cost-effectiveness data across regions. But it is not just absolute unit prices, that are important. Relative prices and other health care system factors can also impact cost-effectiveness or treatment decisions. This broad category comprises of health care resources, characteristics of incentives and providers, which are related to health care organizations and are known to vary across the jurisdictions.

Health care resources

The absolute and relative level of resources available for health care production is different across geographic areas. For example, in 2005 USA dedicated 15,3% of its GDP to health expenditures comparing to Great Britain with 8.4%, 10.6% in Germany, 10.0% in Canada and 9.0% in Italy (OECD, 2008). The absolute level of resources available for health care funding will effect both clinical practice as well as access and dissemination of new diagnostics and treatment alternatives. For instance, cost intensive chemotherapy for treatment of lung cancer produces significant health gain but at substantial additional costs. Funding of the new drugs in less resourced systems will produce higher opportunity costs than in well-resourced systems, since the first will need to withdraw more resources dedicated to treatment of all other conditions to provide for new resource-intensive therapy. The level of opportunity costs of adopting a

new technology may influence political decisions, resulting in rejection of new technology in less resourced systems and adoption in well-resourced. This may be reflected in different willingness-to-pay thresholds for reimbursement between locations on the basis of the essentially the same cost-effectiveness outcomes. As a result, there is higher probability, that cost-intensive diagnostics or therapy will become a standard treatment in well-resourced system comparing to less resourced counties.

Availability of new technology varies across countries. Lower availability (e.g. number of computer tomography per capita, number of laparoscopic surgeries) may lead to inexperienced use; lower case numbers may have effect on costs estimates. The longer distance to technology may effect higher productivity losses associated with treatment. The existence of waiting lists as well as different range of licensed health care products suggest possible variation in technology access across different countries (Welte et al., 2004).

Remuneration and incentives

Health policy and regulations influence relative and absolute prices and set incentives for production and consume in health care. Patients' treatment seeking behavior is influenced by health insurance level with respective taxations, co-payments and deductibles – the factors known to vary across locations. Also, there are direct bonus systems implemented in various health systems to reward health-beneficial behavior. In The Netherlands and in Germany the insurance coverage for dental procedures will be higher, if the patient undergoes regular dental examination. Prices for the same drugs may vary substantially across countries. Price level of drugs is influenced by number of instruments implemented by regulators and payers to control rising drug expenditures, such as positive/negative lists, generic substitutions, co-payments etc. To some extend, the differences in performance of clinical personnel may be explained by incentive structure. Remuneration arrangements based on performance of clinical staff may raise sensitivity to the resource use and total costs and result in shorter duration of hospitalization (Forsberg et al., 2001). Supplier induced demand, i.e. provision of non-vital services, is a known phenomenon in health care. Retrospective remuneration for performed services may indorse supplier induced demand, whereas prospective payments (e.g. based on diagnosis – related groups) may limit it (Welte et al., 2004).

Providers

There is known variation in providers' structure and capacity within and between geographic areas. Number of beds per capita, lengths of stay in hospital, rates of hospitalizations, rates of diagnostic testing, the frequency of clinical follow-ups, and drug dosing regimens are all factors known to vary across jurisdictions. For instance, in 2003 there were 7.59 hospital beds per 1000 population in France, comparing to 8.74 in Germany, 3.45 in Spain and 5.82 in Switzerland (WHO, 2009). Other provider characteristics relate to experience, education, training, skills, efficiency, as well as level of experience with new technology. The skills of clinicians are particularly important for invasive interventions, surgery and diagnostic procedures. However, even non-invasive procedures can display differences between clinicians. The legal obligations of providers may impact clinical decisions. Depending on the disease and evidence around treatment effectiveness, patients may be treated less aggressively in some countries where the threat of legal litigation is more pronounced. This variability may be further enforced by different management of adverse events. These effects are difficult to quantify. Finally, the variation in performance of medical personnel can be due to different incentives in health-care system (Roberts, 1999; Manca & Willan, 2006; Glick et al., 2003; Sculpher & Drummond, 2006a). All these factors can have a significant impact on the effectiveness and cost-effectiveness of a program or service.

3.2.4 Methodological Factors

Study design, conduct and analysis

Health economic studies are often conducted along clinical trials. Randomized controlled studies facilitate assessment of clinical effects in controlled setting using protocolized procedures. Blinding, randomization and standardized procedures improve internal validity of the collected evidence. However, strong internal validity might limit external validity due to protocol-driven procedures, improved compliance, selection of the sites which are not representative for the general population. Also, selected comparators might not be available in all settings. As implication, incremental effectiveness may not be relevant for the countries, where particular comparators are not used.

Other issues affecting generalizability relate to study perspective and measurement of clinical and economic endpoints, which may vary according to different requirements for health economic evaluations. The perspective of the study (payer or societal) can limit generalizability, if the scope used in the analysis does not meet the requirements of the target country. Although, this problem can be usually addressed by disaggregated reporting of study results, not all studies adopt broad societal perspective and publish results in a disaggregated form. Resource use structure varies between the settings and so the estimation within clinical trial setting may not always reflect true costs of technology. Calculation of direct costs can be done using per diem costs, market price weights, listed fees. The costs may account for overhead and capital costs, which could be differently assigned across the clinical institutions even within the same country. Similar to direct costs, estimation of indirect costs, such as productivity impairment, can be measured in several ways. Hourly rates assigned to missed employed work and lost productivity and opportunity costs for unpaid work in household are common estimation methods (Taghreed et al., 2003; Welte et al., 2004; Johannesson & Karlsson, 1997).

Timing

Variability of health economic outcomes over the time has not been extensively studied, partially due to the complexity and funding issues for the studies with the long follow-up period. In general, majority of the "geographic" factors discussed above, can also change over the time. Economic theory suggests that the mean and marginal cost of production may fall with increased volume, if the starting level of production is sufficiently low. In addition, efficiency may be improved with gaining experience with new technology and learning how to optimize production (Ramsay et al., 2001). Experience may be associated with both individual knowledge of clinical personnel, i.e. technical capabilities to apply complex technologies, and common experience of organization with particular instruments, methods and processes (Sculpher et al., 2004). The timing of the study may as well impact cost-effectiveness results. Health economic study conducted at the early stage of the life-cycle of particular technology may not capture the application within routine clinical practice. In addition, clinical practice may change over the time when new therapeutic and diagnostic options become available (Coyle & Lee, 1998). Another time-dependent parameter relates to diagnostics, standard treatment and concomitant therapy, which are likely to change over time and so the

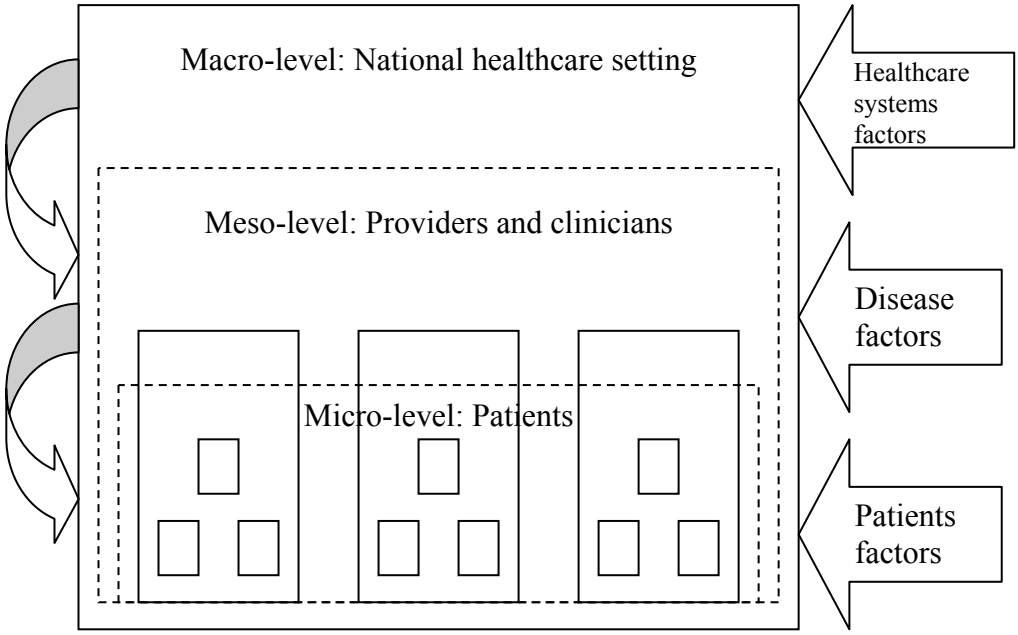
cost-effectiveness results against the alternative technologies for general population or it sub-groups. For instance, evolution of therapies for rheumatoid arthritis is a recent example, where it changed from various combinations of non-steroidal anti-inflammatory drugs (NSAIDs) and analgesics (pain-relieving drugs) to corticosteroids, disease-modifying antirheumatic drugs and biologics, such as anti-TNF. The drugs taken as commonly used alternative treatment in early economic evaluations are probably no longer relevant in current clinical practice (Sculpher et al., 2004).

3.2.5 Complex Cross-Level Interaction Between Variability Factors

Variability is caused by interaction of multiple factors at different levels. Exploring the impact of the variability factors discussed above on the outcomes of different treatments is essential for the development of generalizability methods. Figure 6 provides a simplified picture of operating levels within a health care system with corresponding variability factors. Some factors may have complex interaction on different levels. Thus, the context and the outcomes of intervention should not be studied separately.

The assessment of macro- and micro-level factors requires additional data collection. Some of the factors can be measured directly in the study, e.g. patients age, smoking status, education. The association between these factors and treatment effects can be then explored in the study analysis. Other high-level factors may not be easily captured, e.g. impact of remuneration on treatment pattern and resource utilization. Nevertheless, these effects may be central for health policy decision-making and implementation. Therefore, the acknowledgement and assessment of the effects of different factors is central generalizability question (Murray et al., 2000). Analysis may not be limited only to the evidence which can be easily quantified. Analytical methods should aim to accommodate complex interaction while accounting for heterogeneity and uncertainty in different levels. Following sections describe approaches suggested by international health economic guidelines and published evidence in considering generalizability in design, analysis and results of patient-level studies and decision models.

Figure 6: Hierarchical Levels within Health Care Systems and Variability Factors



Source: compiled by the author

3.3 Generalizability Methods Suggested by Health Economic Guidelines

We reviewed 25 health economic guidelines available in English, German or Russian languages concerned with HTA evaluation and/or submission for reimbursement to national payers and 1 general methodological guideline. The methodological guideline to address generalizability (transferability) in health economic evaluations was developed by ISPOR tasks force in 2008-2009. The recommendations from national guidelines for handling of international evidence and potential methods to improve generalizability are summarized in Table 2.

Table 2: Recommendations of International Health Economic Guidelines for Handling Generalizability

Country	Title of health economic guidelines	Generalizability	Acceptability of international data	Suggested approaches/methods
Poland	Proposal of Polish Guidelines for Conducting Financial Analysis and Their Comparison to Existing Guidance on Budget Impact in Other Countries (Orlewska & Mierzejewski, 2004)	Yes	Yes; international clinical evidence is acceptable.	Systematic review, methods of evidence based medicine.
UK	Guide to the single technology appraisal process (NICE, 2009) Guide to the methods of technology appraisal (NICE, 2008)	Yes	Yes; clinical data may be considered generalizable from international evidence.	Decision-analytical framework for the evidence synthesis.
Belgium	Pharmaco-Economic Guideline (English version) (KCE, 2008)	Yes	Yes	Knock-out criteria to assess whether data is generalizable to Belgium, adjustment to models needed to reflect Belgian clinical and resource use patterns.

Country	Title of health economic guidelines	Generalizability	Acceptability of international data	Suggested approaches/methods
Norway	Norwegian guidelines for pharmacoeconomic analysis in connection with applications for reimbursement (Norwegian Medicine Agency, 2005).	Yes	Yes; international evidence for clinical data may be acceptable, health utility and economic data from outside evidence may not be applicable.	Sensitivity analysis, involvement of local experts to evaluate adjustments to the country-specific context, replacement of abroad resource use with Norwegian data.
Canada	Guidelines for the Economic Evaluation of Health Technologies: Canada (Canadian Agency for Drugs and Technologies in Health, 2006).	Yes	Yes, effectiveness and not efficacy is of primary interest, economic and utility data may not be generalizable.	Homogeneity test to evaluate country-by-treatment interaction, decision models, sensitivity analysis, simple replacement of clinical and economic parameters with country-specific data may not be appropriate.

Country	Title of health economic guidelines	Generalizability	Acceptability of international data	Suggested approaches/methods
New Zealand	Prescription for Pharmacoeconomic Analysis. Methods for Cost-Utility Analysis (PHARMAC, 2007).	Yes	Yes; utility data measured by EQ5D should be calculated using country-specific preferences.	Sensitivity analysis, expert panels for validation of resource use and treatment patterns.
Scotland	Guidance to Manufacturers for Completion of New Product Assessment Form (Scottish Medicines Consortium, 2007).	Yes	Yes; evidence from UK will be in general accepted.	Systematic appraisal of the evidence, meta-analysis and decision modelling
Sweden	Working guidelines for the pharmaceutical reimbursement review (The Swedish Pharmaceutical Benefits Board, 2008).	Yes	Yes	Decision models for evidence synthesis of international and country-specific data.

Country	Title of health economic guidelines	Generalizability	Acceptability of international data	Suggested approaches/methods
France	French Guidelines For The Economic Evaluation Of Health Care Technologies (Collège des Économistes de la Santé, 2004).	Yes	Yes; adaptation for French context required, international clinical data may be used, economic data is less generalizable.	Not specified.
Germany	Entwurf einer Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2009). Allgemeine Methoden (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), 2008).	Yes	Yes; evidence should reflect German health-care setting.	Not specified.

Country	Title of health economic guidelines	Generalizability	Acceptability of international data	Suggested approaches/methods
The Netherlands	Guidelines for pharmacoeconomic research, updated version (College voor zorgverzekeringen Diemen, 2006).	Yes	Yes; utilities from international studies may be used.	Not specified.
Baltic Countries	Baltic Guideline for Economic Evaluation of Pharmaceuticals (Pharmacoeconomic Analysis) (Behmane et al., 2002).	Yes	Yes; data should be adapted to the country-specific context.	Not specified.
Finland	Guideline For Preparing a Health Economic Evaluation (Ministry of Social Affairs and Health, 2003).	Yes	Yes; clinical practice and costs should refer to Finnish context.	Not specified.
Irland	Irish Health care Technology Assessment Guidelines (IPHA/HSE, 2000).	Yes	Yes	Not specified; assumptions made to adjust data to Irish context should be explained.

Country	Title of health economic guidelines	Generalizability	Acceptability of international data	Suggested approaches/methods
Italy	Guidelines For Economic Evaluations in Italy: Recommendations From The Italian Group Of Pharmacoeconomic Studies (Capri et al., 2001).	Yes	Yes	Not specified.
Portugal	Guidelines For Economic Drug Evaluation Studies (Infarmed, 1998).	Yes	Yes	Not specified.
US	The AMCP Format for Formulary Submissions (AMCP, 2005).	Yes	In general yes; the manufacturer must demonstrate that international data is applicable for US setting.	Not specified.

Country	Title of health economic guidelines	Generalizability	Acceptability of international data	Suggested approaches/methods
Austria	Guidelines on Health Economic Evaluation (Walter & Zehetmayr, 2006).	Yes	Yes; evidence from EU countries and Switzerland may be considered applicable to Austrian setting.	Not specified.
Hungary	Methodological guidelines for conducting economic evaluation of healthcare interventions in Hungary: a Hungarian proposal for methodology standards (Szende et al., 2002).	Yes	In general yes, if relevance for Hungarian setting is demonstrated.	Sensitivity analysis.
Russia	Branch Standard. The Standardization System in the Russian Federation Health Care System Clinico-Economic Studies. General Provisions (RF Ministry of Health, 2002).	No	NA	NA

Country	Title of health economic guidelines	Generalizability	Acceptability of international data	Suggested approaches/methods
Switzerland	Kriterien zur Beurteilung des Nutzens von komplementärmedizinischen Methoden (Heusser, 2001).	No	NA	NA
Denmark	The Danish approach to standards for economic evaluation methodologies (Alban et al., 1997).	No	NA	NA

Source: International Guidelines, ISPOR "Pharmacoeconomic Guidelines Around the World" database (ISPOR, 2010), compiled by the author.

Recommendations of international health economic guidelines

Overall, international health economic guidelines differ in their perspectives, requirements and recommendations related to generalizability of international evidence (Barbieri et al., 2005; Drummond et al., 2009). These differences may be partially explained by various roles and objectives pursued by national HTA bodies, as well as access and availability of country-specific evidence. The majority of reviewed health economic guidelines acknowledge generalizability issues related to the data collected in multinational studies and the differences in the clinical and economic parameters between and within the countries. Only 3 out of 25 guidelines are not directly referring to the generalizability problematic (what does not mean that decision-makers in these particular countries will not require this information with submission documents). Guidelines differ in scope and extend to which they discuss approaches to handle international data. In general, there is tendency to accept clinical evidence to the greater extend than the economic data from the abroad settings. The variation of health preferences and utility across geographic areas is acknowledged, however, there is no common convention about acceptability of health utility data collected outside the setting of interest. Majority of guidelines require use of country-level economic data or it adaptation to reflect country-specific resource use and costs. However, the recommendations about potential approaches and data sources vary across guidelines. Selected guidelines reference to the national costs and unit prices. 8 out of 25 guidelines make suggestions about the potential methods to improve generalizability. These methods are mainly related to systematic evidence appraisal and evidence synthesis by use of decision analytical framework. Canadian guidelines suggests homogeneity test to check country-by-treatment interaction for the evidence collected in multinational trial.

Recommendations of ISPOR transferability guideline

The methodological guideline to address generalizability (transferability) in health economic evaluations was elaborated by ISPOR tasks force in 2008-2009. This guideline provides an overview of available analytical tools to improve generalizability in patient-level studies and decision models. In essence, while addressing and dealing with generalizability in patient-level studies, ISPOR guideline suggests to explore heterogeneity by using descriptive statistics, heterogeneity test for qualitative and quantitative interactions. Hierarchical models are recommended to account for

clustering within sites or countries. The generalizability approaches for decision models are largely depending on availability of setting-specific data and the clinical course of the condition. Several approaches are discussed, including construction of event-based models, application of trial-wide efficacy estimates to the country-specific baseline case-mix, synthesis of clinical evidence from randomized-controlled studies through meta-analytical approach. The input data for clinical and economic parameters in a model can be estimated based on the results from one trial from the target country or on a synthesis or meta-analysis of relative treatment effects collected in several clinical trials. Adjustment of data collected in international trials to the country-specific baseline parameters (case mix) can be supported by regression modelling with baseline data as covariates. Extensive deterministic and probabilistic sensitivity analysis is the key tool to explore sensitivity of results to different assumptions related to generalizability of the data (Drummond et al., 2009).

3.4 Generalizability Methods for Decision Models

Decision models and patient-level studies are complementary approaches for the standard health economic analysis and for generalizability issue specifically. Decision modelling could substantiate cost-effectiveness analysis in the following cases:

- Patient-level study was not conducted in target county and available (patient-level) data may not be fully relevant and applicable for the specific setting.
- Availability of additional relevant data sources in the target country, e.g. observational clinical and economic evidence, registries, HRQoL data, epidemiological data on baseline risks, etc.
- Data collected in patient-level studies do not fully meet the requirements of health economic analysis, e.g. insufficient follow-up period, protocol-driven procedures or irrelevant treatment comparator for the target country.
- Available decision model synthesizing evidence from different sources can be updated with the respective information from a target country (Sculpher et al., 2006b; Drummond et al., 2009).

Technically, modelling strategy can include formalized approaches such as decision trees, Markov analysis, discrete event simulation and systems dynamics. Practically, the choice of the modelling strategy depends on specific aspects of clinical condition, health economic impact of a new treatment and, finally, access and availability of the data (Brennan & Akehurst, 2000).

The major advantage of decision modelling is the facilitation of the data synthesis from different sources, in particular for the cases, where clinical studies capture incomplete information (Urdahl et al., 2006). Decision models provide the framework to apply generalizability methods to construct cost-effectiveness results for the long-term therapies, extent results to non-trial locations and to adjust trial results to reflect routine practice. The individual aspects are explained in the following sections.

3.4.1 Modelling Long-term Results

Decision-models allow simulation of long-term costs and benefits based on data collected in short-term clinical studies and assumptions or epidemiological evidence about clinical progression and treatment effects. The transition probabilities for Markov models can be taken from epidemiological data or observational studies. The economic data on costs and resource use can be collected within local costs studies or found in existing national registers.

Generation of long-term results beyond clinical study frame includes:

- prediction of real-life effectiveness based on short term clinical efficacy
- estimation of total (life-long) resource use and discounted cost (Brennan & Akehurst, 2000).

The advantage of combining different type of data could be potentially source for error. The absence of randomization in observational data raises concern about selection bias. A statistical method to deal with not randomized data - “decomposition technique”- has been applied by Shih and colleagues in the context of model for treatment of patient with renal disease with epoetin (Shih & Kauf, 1999).

Decision models are used for number of chronic conditions. For example, Kobelt and colleagues (2003) constructed cost-utility model for interferon treatment of patients with

multiple sclerosis (MS). Model synthesized data from different sources, including clinical studies and public register for MS progression. Analysis estimated the outcomes of treatment over 40 cycles (10 years) based on the data from clinical study for 12 cycles (3 years). Ordered probit regression was applied to model transition probabilities using data from clinical trial. Disease stage was used as dependent variable and age at onset of MS, age at start of treatment, time since start of treatment, relapse in the current period, and six dummy variables representing the state in the previous period as independent variables (Kobelt et al., 2003).

3.4.2 Modelling Location-Specific Results

Decision models provide location-specific economic evidence for both study and non-study settings. Where patient-level data is available, but the targeted country was not included into the study, Manca and Willan (2006) suggested to develop an "event-based" model. Their framework is based on the generalizable aspects of clinical condition and patient-level data from the available studies. Country-level estimates can be modeled in applying pooled incremental efficacy to the location-specific baseline parameters. Application of pooled incremental efficacy (e.g. reduction of pain, adverse events) to the baseline parameters in specific countries is a common and feasible approach, if there is no evidence to suggest that use of relative efficacy from a specific setting or country is more appropriate (Manca & Willan, 2006, Drummond et al., 2009).

Caro and colleagues (2000) used similar methodology to set-up a generalized cost-effectiveness formula distinguishing between the parameters, which could be considered as general or country-specific. This approach supported generalization of results from a clinical study in cardio-prevention conducted in Scotland to Belgian setting. Incremental efficacy was considered generalizable, i.e. applicable to all countries. The baseline parameters were taken from country-specific sources. Estimated relative risk reduction in Scottish population was applied to Belgian-specific baseline data and costs. The cost-effectiveness ratio for Scottish and Belgian setting were not substantially different. The authors concluded, that results could be valid for other countries with comparable clinical characteristics such as cholesterol level, age and gender (Caro et al., 1999; 2000; Sculpher et al., 2004).

Availability and access to country-specific baseline clinical evidence is critical for the majority modelling approaches, including the discussed above. Statistical models are often used in health economic analysis to estimate future events in selected clinical indications, such as fractures in osteoporosis or infarction in cardiovascular indications. For example, Framingham cardiovascular risk equations were developed and populated with US data. In fact, the logic of this algorithm may be applicable to the other countries with the comparable clinical practice and patient population. However, the update with country-specific baseline data to reflect local case-mix is always commendable (Drummond et al., 2009).

For modelling of health economic outcomes when the patient-level data is not available and particular country did not participate in the trial, Manca and Willan (2006) outline the major issues and reviewed these in a case study conducted by Palmer and colleagues (2005) to evaluate cost-effectiveness of glycoprotein IIb/IIIa antagonist in treatment of patients with non-ST-elevation acute coronary syndrome. Critical analytical and empirical issues included adjustments needed to reflect country-specific clinical practice, comparison of the clinical baseline data with country-specific case-mix, evaluation of the relationship between baseline risk and clinical efficacy of the particular therapy and, finally, application of country-specific economic parameters (Manca & Willan, 2006; Palmer et al., 2005). The "efficiency" of the analysis requires maximum utilization of available evidence under reflection of targeted context. For this purpose, decision models offer advantage of combining data from different sources. The clinical effectiveness of the treatment (for e.g. calculation of transition probabilities in Markov model) can be taken from clinical trial. In the case there are multiple sources for the parameters of interest, the data should be pooled using appropriate techniques to reflect imprecision of the estimated outcomes and possible heterogeneity. Meta-analytical techniques have made progress to allow synthesis and handling of different types of outcomes from different sources (Drummond et al., 2005).

There are several issues with modelling of health economic outcomes for the settings, where clinical or economic data is scarce and relies on the outside evidence. First, the health production function must be the same as in the study setting, which is rarely the case, considering different source of variability discussed above. Second issue is the potential source for error while applying non-randomized observational data. Finally, economical theory suggests the joint consideration of resource use and costs. Simple

application of the country-specific unit costs to external resource use, is not always appropriate (Reed et al., 2005; Manca & Willan, 2006).

3.4.3 Modelling "Routine Clinical Practice" Results

Clinical trials are often done in the sites, which are specialized on doing clinical research. In general, if the decision model is build on the efficacy data from clinical trial, the adjustment to routine clinical practice should be considered.

Clinical trials are known to raise better treatment compliance, comparing to the everyday practice. This issue should be explored in the sensitivity analysis to show the impact on the cost-effectiveness ratio.

The characteristics of clinical condition and health technology have major implications for the design of health economic evaluations. Until now, the majority of pivotal clinical studies are being done against placebo to meet regulatory requirements for marketing authorization. For many clinical conditions control arm with placebo treatment may not be feasible and/or ethical to implement. Models allow to incorporate all treatment options with respective sequences to reflect current clinical practice to inform policy in a specific setting (Brennan & Akehurst, 2000). The generalizability of paclitaxel study by Beard and colleagues (1997) required incorporation of relevant standard treatment in UK. The trial was conducted against cisplatin and cyclophosphamide, whereas carboplatin was used as standard therapy in UK. The results showed improved cost-effectiveness of paclitaxel vs standard UK treatment comparing to the control arm used in clinical study (Beard et al., 1997).

Baltussen and colleagues (1999) proposed assessment and adaptation of study results to reflect routine clinical practice using three-step approach. They evaluated internal and external validity of the study outcomes and estimated net impact at system level. The assessment of internal validity relied on analysis of study-specific results, which is the first step of the proposed approach. In the second step, authors adjusted study-specific results in accounting for country-level factors related to medical staff, hospital and health care system. This approach involved qualitative assessment, inclusion of epidemiological evidence and sensitivity analysis. In the third step, context-specific

clinical outcomes and costs were evaluated to estimate net impact of new intervention (Baltussen et al., 1999; Sculpher et al., 2004).

3.5 Generalizability Methods for Patient-Level Studies

Conduct and analysis of multinational patient-level studies is a complex and lengthy process. Generalizability can be improved, if the issue is considered already on study design phase, appropriately incorporated in analytical strategy and, finally, published in the form useful for external settings. We discuss respective considerations for design, analysis and report in the following sections.

3.5.1 Considerations for the Design of Patient-Level Studies

Economic evaluations based on patient-level data collected alongside randomized controlled clinical studies continue to provide an important source of data for cost-effectiveness analysis (Drummond et al., 2005). Multinational clinical studies are considered as "state of the art" for producing estimates on clinical effectiveness and safety. The main motivation for conducting of the studies internationally is the large sample size, rapid recruitment and collection of data in wide range of treatment settings across different countries. However, interpretation and generalizability of the results from international studies is problematic for both clinical and economic outcomes (Cook et al., 2003). Additionally to the variability factors discussed in the previous sections, economic evaluations within clinical trials are problematic, as sample size of clinical studies is often determined to show the differences for clinical and not economic end-points and number of patients in different settings is often small and imbalanced. In addition, at the early development stage, the resource use in "routine practice" for new technology can not be easily reflected within clinical trial. Missing, censored or non-existing data is another important issue (Greiner et al., 2000; Drummond et al., 2005). Design of clinical studies often introduces protocol-driven costs, involves very selective procedure for patient sample recruitment, which may not be representative for the general population, and have placebo-arm or comparators not reflecting routine clinical practice (Sullivan et al., 2001).

Considering generalizability issue at the study design stage, it is important to create a framework, where valid economic patient-level data from the representative population can be collected. For the purpose of generalizability, inclusion of study centers would be ideally done randomly from the geographic regions for which economic data is requested. Inclusion and exclusion criteria for the study patients should avoid selecting a specific sub-group not representative for underlying patient population. Important to consider, that the average case load can differ across locations. E.g. centers of excellence or university clinics often treat patients with advanced disease stage or those with severe condition, which may not be the average case for ambulatory practice. In this case, it is crucial to collect both patient-level and center-level variables (Drummond et al., 2005; Goeree et al., 1999).

Comparative treatment should represent standard clinical practice for the jurisdiction, for which reimbursement decision will be made. Today, it is rarely the case, that standard of care does not exist and placebo control is justifiable for treatment control arm. Nevertheless, generalizability could be problematic, if standard treatment differs across jurisdictions (Drummond et al., 2005).

To generalize results of clinical trials, it is important to reflect routine clinical practice and average case-mix of treated population. This is an issue in clinical trials, where enrollment criteria exclude part of patient population, which is going to receive treatment once technology is licensed. Thus, inclusion criteria should be either “broad”, or analysis should be substantiated by additional observational study. Protocol-driven costs and utilization could lead to over- or under- estimation of resource use in everyday practice (Sullivan et al., 2001). While protocol-driven resource utilization may be corrected or subtracted in the analysis, increased compliance in clinical study setting may inflate efficacy estimates, which may not replicate effectiveness in routine practice (Goodman & Mills, 1999). This issue could be explored in the sensitivity analysis using decision model. Observational study design could improve external validity of the clinical endpoints and control for protocol driven resource consumption, but may, on the other hand, fail to provide clinical evidence with maximum internal validity, as required for dug approval purposes (Coyle & Lee, 1998; Tunis et al., 2000).

The important advantage of the patient-level studies is the possibility to collect resource use and cost data at the early stage before the market approval. It is advisable to collect resource utilization in terms of e.g. length of hospitalization, community nurse

visits, physiotherapy etc. separately from the price weights of those resources. This approach has several advantages. For example, the decision-makers in other jurisdictions may need to use local prices to value resource use. Certainly, decision-makers evaluating analysis based on outside evidence, need to assess, whether the clinical procedures and corresponding utilization of resources observed in other locations reflect medical practice in their jurisdiction (Drummond et al., 2005; Brown et al., 2004).

3.5.2 Methods for the Analysis of Patient-Level Studies

Approaches to analyze patient-level data collected in multilocational trials could be divided into fully pooled clinical estimates with multi- or one-country costing, fully splitted clinical estimates with country-level costing, partially split of clinical and economic estimates. Table 3 qualitatively summarizes advantages and disadvantages of these methodological approaches (Reed et al., 2005).

Table 3: Advantages and Limitations of Analytical Strategies for Patient-Level Studies

Advantage of analytical strategy	Fully pooled, multicountry costing	Fully pooled, one-country costing	Partially split, one-country costing	Fully split, one-country costing
Maintain patient-level relationship between resource use and clinical effect	Yes	Yes	No	Yes
Maintain patient-level relationship between resource use and costs	Yes	No	Yes	Yes
Maximize statistical power for treatment effect	Yes	Yes	No	No
Minimize collection of unit costs	No	Yes	Yes	Yes
Allows consistent reporting of treatment effect	Yes	Yes	Yes	No

Source: modified from Reed et al., 2005.

Applying pooled efficacy to resource use with one- or multi-country costing

A rather common approach in multinational trials is calculation of the total costs by applying unit costs from one country or centre to the pooled resource use collected across all countries or centers. Another similar approach is to apply unit costs to resource use data collected in each country and relate pooled costs across all countries to trial-wide efficacy. The feasibility of this approach requires both clinical efficacy and resource use to be fully exchangeable across countries and centers. This assumption

may not be universally valid. Economic theory suggests that clinical institutions, similar to commercial organizations, will attempt to achieve efficient allocation of available resources. Technical efficiency of production function relates to maximum output given the mix of constrained input factors. Clinical sites, being cost-sensitive organizations, would choose the least cost-input allocation to produce targeted outcomes at the level of technical and productive efficiency (Sculpher et al., 2004; Raikuou et al., 2000). In fact, it may not be appropriate to apply price weights from outside jurisdictions.

The implication of separate analysis of costs and resource use will depend on the differences in absolute level of costs and resource use across settings. For instance, application of price weights from US or Western European countries to the resource use in developing countries will inflate both total and incremental costs between the treatment groups. The difference between the costs of new treatment and comparator is the numerator of ICER, overestimation of incremental costs will provide incorrect ICER (Reed et al., 2005). The resource-mix and health production technology are likely to vary across geographical locations and institutions. For example, delivering babies by midwives versus family doctors, and care of non severely sick infants in neonatal intensive care units versus regular nurseries with intensive nursing, are examples of ways of providing a relatively similar service at different costs (Goeree et al., 1999).

A fully pooled analysis with multi-country costing preserves the patient-level relationship between resource use and clinical efficacy and maximizes statistical power. Possible limitation is the need to collect unit costs and make assumption about costing methods in different countries. The methodological approach and assumption in costing may not be completely transparent and standardized across all participated countries. Other limitation relates to possible deterioration of trial-wide difference in resource use if applying country-level prices, since countries with higher price weights may affect ICER outcomes to greater extend than countries with lower prices. Lastly, the application of the pooled results from multinational setting may not be evident and easily interpretable for the decision-makers in participating countries (Reed et al., 2005; Koopmanschap et al., 2001).

Fully split analysis in applying country-level costs to country-level efficacy

The opposite to the fully pooled analysis is the assumption that the data and results from different countries are distinctive and completely not exchangeable. In this case,

the analyst applies unit cost to resource use in relation to clinical effectiveness for the patient sub-group from a specific country only. This approach may be feasible for large countries, like US, Canada and UK, were large-scale national studies are conducted routinely. The statistical power and selection bias may be, however, an issue for the cases where recruitment in particular countries was slow and the analysis should rely on the data from very small patient population (Sculpher et al., 2004; Reed et al., 2005).

Partial pooling of either efficacy, resource use and costs

Partial pooling is probably the most common way of analysis of multinational studies and relies on partial aggregation of either economic or clinical data (Li, 2007). For example, cost estimation may be limited to the data collected in one country, whereas efficacy is estimated from trial-wide patient population. This approach may be considered as a compromise between fully pooled and fully split strategy, i.e. between statistical power of clinical findings and limited generalizability of economic parameters across different countries. Partial pooling allow to maintain the relationship between unit costs and resource use, but not between the costs and clinical effectiveness (Reed et al., 2005). In fact, in many studies partial exchangeability of absolute and relative clinical and economic outcomes can be assumed. The feasibility of combining health economic data across the jurisdiction remains an analytical issue (Cook et al., 2003). Homogeneity test and several regression-based methods are statistical approaches supporting implementation of partial pooling strategy.

Homogeneity test

Cook and colleagues (2003) tested homogeneity of the economic data in the Scandinavian Simvastatin Survival Study. As noted before, it is often assumed, that relative clinical efficacy of the treatment does not vary between different settings. As a result, measures of clinical effect may be aggregated across centres and countries. Usually, homogeneity of treatment effect is checked before the clinical data are pooled. However, detection of statistically significant difference, i.e. country by treatment interactions, is unlikely, partially due to high number of participated sites/countries with small number of patients per setting. In addition, absence of country by treatment interaction in incremental efficacy does not mean homogeneity of absolute clinical outcomes, which are used to estimate cost-effectiveness of new therapy in the denominator in the cost-effectiveness ratio.

Homogeneity test by Gail and Simon (1985) is suggested by researches for testing of economical outcomes similar to the test for investigation of country-by-treatment interaction in clinical endpoints. Treatment outcomes can be measured in terms of incremental resource use, incremental costs, cost-effectiveness ratio and net monetary benefit. Country-by-treatment interactions can be of qualitative or quantitative nature. Qualitative interaction refers to the different "direction" in treatment effect across the settings, e.g. new treatment is saving costs in some countries and adding in others. Quantitative interaction characterizes differences in the effect's magnitude. If homogeneity of effect's direction and magnitude is confirmed by the tests, researcher may consider to aggregate observations across the settings to maximize statistical power and consider trial-wide effect applicable for the decision in individual countries (Cook et al., 2003). Nevertheless, the appropriateness of combining the outcomes across the countries may be challenged by the unbalanced recruitment and small sample size from some countries. The application of this method is demonstrated in the case study and described in details in section 3.6.

Fixed-effect regression models

Number of studies employed regression methods to evaluate generalizability of the trial-wide cost-effectiveness results to individual countries participating in multinational clinical trial. Wilke and colleagues (1998) proposed regression-based method for the trial of tirilazad mesylate treatment for subarachnoid hemorrhage for the five highest-enrolling countries to estimate country-specific incremental costs, while controlling for clinical endpoints and country-specific differences in costs for outcome. The method assumed that the costs of treatment in multi-country model are determined by covariates, such as treatment, clinical outcome, country-by-treatment and country-by-outcome interaction terms. Although country-specific effects on survival were not statistically different, the authors concluded, that generalizability of trial-wide costs to the specific countries is not appropriate, as the greatest variation was observed in using country-specific resource utilization, unit prices and outcome levels. Thus, country-specific cost-effectiveness estimates were calculated with trial-wide incremental efficacy, but country specific incremental costs (Wilke et al., 1998; Wilke, 2003).

Rutten-Van Molken and colleagues (1998) used ordinary least-squares regression to explore heterogeneity in the cost in the cost-effectiveness analysis of two asthma

treatments in the study conducted in six European countries. Country-specific unit costs were applied to resource use of country-level patients subgroups. Regression analysis was conducted using log transformed total costs as dependent variable and country, treatment, age and duration of condition as covariates. Analysis showed no evidence for statistically significant difference in cost between the treatment arms, but significant variation in costs across the countries. Patients from Italy had significantly lower costs, comparing to other countries, whereas patients from Switzerland had the highest costs. This heterogeneity was partially explained by variations in costing approaches and uncertainty about the data sources. The authors emphasized, that pooling of costs may not be informative and suggested splitted reporting of resource use and costs, instead of total costs only (Rutten van Molken et al., 1998; Sculpher et al., 2004).

Coyle and Drummond (2001) applied one-way ANOVA in economic analysis of radiotherapy in cancer treatment to explain heterogeneity in radiotherapy and relevant inpatient costs between and within treatment centers. Variability in costs may arise on patient- and center- levels. Regression analysis was used to identify statistically significant difference in costs between the treatment centers and specific parameters (covariates) which could explain this variation. Identifying the predictors allows to improve the generalizability of the study findings beyond the study settings to e.g. hospitals with the same level of covariates. Analysis revealed significant heterogeneity in costs across the centers. Major part of variation in radiotherapy costs was explained by differences in price weights. The significant differences in inpatient costs were partially attributed to variation in costs for ward's accommodation. Finally, patient-level covariates explained differences in costs between two treatments (Coyle and Drummond, 2001; Sculpher et al., 2004).

Hierarchical regression models

Several authors suggested using hierarchical regression-models to handle multilevel data in cost-effectiveness analysis (Rice & Jones, 1997; Manca et al., 2005, Sculpher et al., 2004; Drummond et al., 2005, Duncan et al., 1998). Observations in multinational studies may be considered as "clustered" within locations, geographical areas or clinical institutions. Stochastic cost-effectiveness analysis, however, relies on independence of the events within centres, countries or geographic regions. Thus, the application of e.g. ordinary least square regression does not formally allow hierarchical (clustered) nature

of the data and will produce inefficient parameter estimates. Hierarchical modelling can be used for clustered data to calculate (1) correct estimates of the trial-wide incremental cost-effectiveness with respective standard errors and confidence or credibility intervals, and (2) country-level estimates of incremental cost-effectiveness for either informing decisions based on study results or for further analysis as input in decision models. The estimation of the statistical models can be done with frequentist's maximum likelihood or Bayesian shrinkage estimation procedures. The caveat of the use of hierarchical models is the assumption, that selection of the sites in the study is random, which is rarely the case (Sculpher et al., 2004; Manca et al., 2005). The detailed discussion and application of this method is demonstrated in the case study and described in details in section 3.6.

Value of perfect information analysis to guide further research

Obviously, generalizability of health economic evaluations is always subject to uncertainty associated with input parameters and results. Decision models allow to synthesize available evidence to establish cost-effectiveness of a new intervention to the general population or it sub-groups and explicitly account for uncertainty associated with estimated parameters. The decision-makers will face inevitable question about the likelihood to make wrong decision based on the current uncertain evidence.

Several advanced statistical methods are being developed in this field. Value of perfect information analysis is a new approach rooted in decision theory. It aimed to support decision about the need for additional studies by assessment of the probability and the costs of making the wrong decision on the basis of current evidence. The intervention should be reimbursed based on available data at a give time-point. Nevertheless, the quantification of the degree of uncertainty provides decision-makers and researches with the valuable information. It can help to determine future research needs, priorities and optimize research design (Claxton et al., 2002; Drummond et al., 2005).

3.5.3 Appraisal Methods for Reported Evidence

Extensive number of publications exist on the assessment and appraising of existing economic evaluations. Several authors suggested various criteria or checklists, which can be used for the assessment of published evidence for the decision-making in particular country (Boulenger et al., 2005; Urdahl et al., 2006; Spath et al., 1999; Welte

et al., 2004). The short list for quick assessment is provided by Drummond and colleagues (2005) in Figure 7. Overall, the authors agree that the essential prerequisite for the utilization or generalizability of "external" data is the solid internal validity of the evaluations and availability of the detailed information on effectiveness, costs and resource use.

Figure 7: Checklist for Assessment of the Generalizability of Published Studies

1. Are study sites representative of the jurisdiction(s) for which data are required?

2. Are study sites (centers) randomly selected?

3. Can data on center characteristics be collected (e.g., bed occupancy levels)?

4. Does the trial include a high proportion of the normal clinical caseload?

5. Does the comparator therapy represent current practice in the settings concerned?

6. Is a wide range of user perspectives represented in the study?

7. Are prices (unit costs) being collected separately from resource use data?

8. Is a widely used generic instrument being used for quality of life measurement?

9. Can regression-based techniques be used to obtain center-specific measures of cost-effectiveness?

Source: modified from Drummond et al., 2005.

3.6 Methods Applied In Patient-Level Case Study

Selection of analytic strategy to explore variability and improve generalizability depends on several factors. The essential factor is the availability of the patient-level data. For the patient-level studies, two methods were suggested by the literature: homogeneity test and hierarchical modelling. First method is deemed to detect heterogeneity in the incremental effects and the second to calculate country-level estimates.

Our case study is based on secondary analysis of patient-level data collected in clinical phase III study comparing efficacy of two antibiotics in treatment of community-acquired pneumonia (CAP). The detailed description of study design and

dataset is provided in section 4. This section presents methods applied to improve generalizability of clinical and economic results. We applied the selected generalizability methods on the following endpoints:

- Mean and incremental effectiveness.

Mean effectiveness is measured by proportion of cured patients in the treatment and comparator arms. Incremental effectiveness is the difference between treatment and comparator arms. Positive incremental effectiveness means that treatment is more efficacious than comparator.

- Mean and incremental costs.

The costs outcomes are measured similarly to the effectiveness. Positive incremental cost suggests that treatment is more expensive than the comparator therapy.

- Mean and incremental length of stay (LOS) in the hospital.

Hospitalization is the major resource use in treatment of CAP. More broadly, in-patient care contributes substantially to the overall costs in health care sector. The between-country heterogeneity and impact of the new therapy on LOS is therefore of particular interest and is being evaluated.

Details on homogeneity test and hierarchical models applied in the case study to the above endpoints are described below.

3.6.1 Homogeneity Test

The test was suggested by Cook and colleagues in analysis of multinational health economic study to explore homogeneity of the effect in different settings (Cook et al., 2003). The test evaluates, whether the relevant clinical or economic effect differs by country and treatment. The treatment effect D_k with the variance s_k^2 for countries k=1 to K can be an estimate for difference in clinical response between treatment and comparator arms, difference in length of hospitalization, difference in costs, incremental net monetary benefit (INMB). The between-country heterogeneity can be of qualitative or quantitative nature. Qualitative interaction refers to the differences in the effect direction, whereas quantitative interaction refers to the effect size or magnitude. If there

is no evidence of heterogeneity and sample size was sufficient to detect economically meaningful differences in effectiveness and costs, then the aggregated trial-wide estimates are considered valid for all participated countries. If the method indicates heterogeneity, the trial-wide estimates may not be informative for the decision-making in individual countries (Cook et al., 2003; Gail & Simon, 1985).

Test for qualitative interactions

Qualitative interaction refers to the heterogeneity where the treatment effect (e.g. difference in clinical effectiveness, resource use, costs, net monetary benefits) is positive in some countries and negative in the others. Finding of qualitative interaction suggests that e.g. treatment is more efficacious in some countries, whereas comparator is more efficacious in the others, or new therapy has positive NMB in some countries and negative in the others.

In the case study, we test the null hypothesis that there is no country-by-treatment interaction, i.e. the treatment effect is either greater than 0 or less than 0 for all countries as following:

1) $\Delta E_k \geq 0$ in all countries. Incremental effectiveness $\Delta E_k = TOC_{Tk} - TOC_{Sk}$, measured as difference in treatment response, is greater in the treatment arm comparing to the standard therapy.

2) $\Delta C_k \leq 0$ in all countries. The treatment is cost saving comparing to standard treatment in all countries, i.e. incremental cost $\Delta C_k = C_{Tk} - C_{Sk}$, measured as difference between mean total costs for treatment and comparative therapy, is non-positive for all countries.

3) $INMB_k \geq 0$ in all countries. Incremental monetary benefit calculated as $INMB = \Delta C\lambda - \Delta E$, is non-negative in all countries.

4) $LOS_{Sk} - LOS_{Tk} \geq 0$ in all countries. The mean length of stay for the patients receiving comparative therapy is longer than for those getting new treatment.

The test statistic is computed as following:

$$Q^{-} = \sum_{k=1}^{K} (D_k^2 / S_k^2) \text{ for } D_k \geq 0$$

$$Q^{+} = \sum_{k=1}^{K} (D_k^2 / S_k^2) \text{ for } D_k < 0$$

The likelihood ratio test is expressed by:

$$Q = \min(Q^{+}, Q^{-}) > C_{2\alpha}$$

Critical values are reported in Gail and Simon (1985).

The calculation of the power of the test for qualitative interactions is suggested by Pan and Wolfe (1997) as following:

$$1 - \prod_{k=1}^{K} \Phi[D_k / S_k + z_\alpha] - \prod_{k=1}^{K} \Phi[-D_k / S_k + z_\alpha] + \prod_{k=1}^{K} \{\Phi[D_k / S_k + z_\alpha] - \Phi[-D_k / S_k - z_\alpha]\}$$

Where $\Phi[\ ]$ is the standard normal cumulative distribution function and z_α is the α percentile of the standard normal distribution for $\alpha = 1 - (1 - \alpha)^{[1/(k-1)]}$.

Test for quantitative interactions

This test evaluates homogeneity of treatment effect's magnitude across different settings. The null hypothesis for the treatment effects (as formulated for qualitative homogeneity test) is that the effect's magnitude is equal in all countries.

Test statistic is computed as following:

$$H = \sum_{k=1}^{K} (D_k - \overline{D})^2 / S_k^2$$

$$\overline{D} = \left[\sum_{k=1}^{K} D_k / s_k^2\right] / \left[\sum_{k=1}^{K} 1 / s_k^2\right]$$

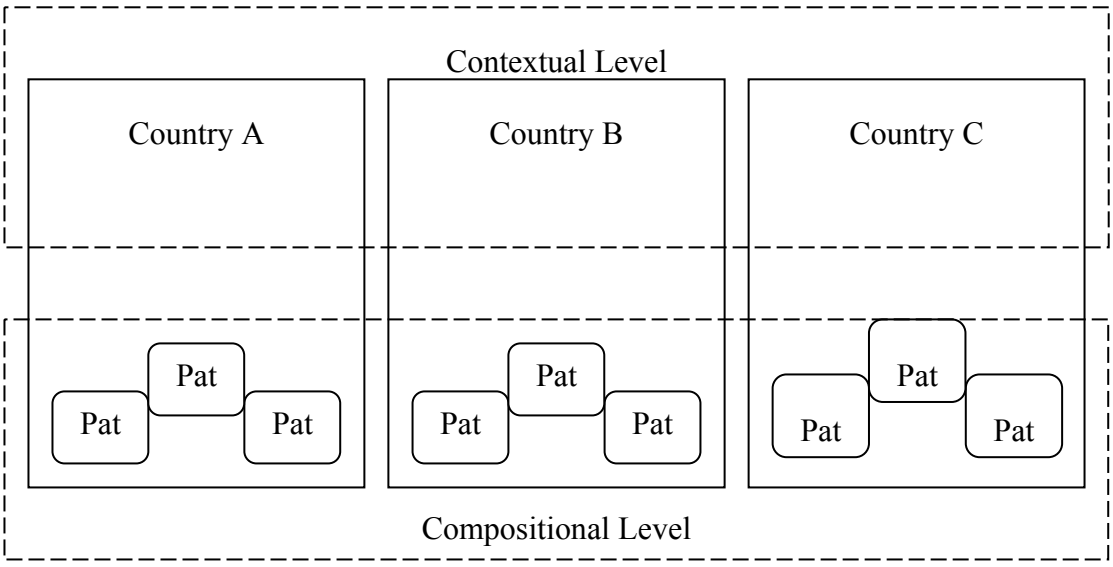
Test statistics H is compared with critical values of the chi-square distribution with K−1 degrees of freedom. Heterogeneity is indicated by large values of H, where treatment effect in individual countries departs substantially from trail-wide estimate.

3.6.2 Hierarchical Modelling

Various research areas in social science are evaluating associations between individual factors and environmental context. This question is important for Public Health research where several studies explored relationship between contextual impact and individual health status (Rice & Jones 1997; Duncan et al., 1998; Britton, 1990).

Variations between different countries and sites in clinical and economic parameters in the health economic study can be associated with both clinical characteristics of the patients in the setting ("compositional" effects), as well as the factors of the environment where treatment is provided ("contextual" effects), as illustrated in the Figure 8. The outcomes under study may be affected by processes operating at more than one "level": patient-, centre-, health care system- levels. To understand and quantify contextual and compositional impact on cost-effectiveness outcomes, all relevant levels need to be incorporated into analysis and considered simultaneously (Duncan et al., 1998).

Figure 8: Contextual and Compositional Levels in Clinical Study Setting



Source: compiled by the author

Traditional methods for analysis of contextual and compositional effects

Traditional regression analysis have limitations in analyzing contextual and compositional levels together. Standard regression models, estimated with ordinary least-square approach, allow for one-level analysis, e.g. aggregated (contextual) level or

individual (compositional) level. The issue with aggregated analysis is known under the term of “ecological fallacy”, where uncovered macro-level inferences may not reflect lower level associations. The disadvantage of conducting analysis at a lower level is potential “atomistic fallacy” caused by neglecting of environmental impact where the treatment and health outcomes are being produced (clinical trial site, in-patient institution, country) (Alker, 1969). Clustering imposes the similarity or dependency of observations coming from the same higher-level unit, what will produce inefficient estimates and lead to inflation of type I error, e.g. finding differences and associations in the cases where these may not exist (Skinner et al., 1989; Snijders & Bosker, 1999).

ANCOVA analysis allows to accommodate hierarchy in introducing additional “fixed”-effect as proxy for different sites or countries. For example, the analysis over 50 settings (clinical hospitals) with different intercept and slope terms for each would result in 100 parameters. This, in turn, possess serious requirements for the sample size to assure the reliability and robustness of the estimates. Thus, ANCOVA has limited capability to maximize efficiency in the analysis of small dataset. Another limitation of the “fixed”-effect approach is the fact that the inferences are valid explicitly for studied sample and context, but not necessarily to general population, what is of main interest in Public Health research. Instead, hierarchical models make inferences for the underlying population, from which the studied sample was randomly drawn. Overall, hierarchical modelling overcome the issues related to aggregated and disaggregated analysis and aid to fully explore the context and improve generalizability of the estimated parameters (Duncan et al., 1998).

Application of hierarchical models for clustered data

Hierarchical models are fairly recent statistic development. Different termini are used in the literature to describe this methodology: multilevel models (Goldstein, 1995), random coefficient models (Longford, 1993) and hierarchical models (Bryk & Raudenbush, 1992). In this work we use the term "hierarchical models". Hierarchical models are applied in various research areas of social and natural sciences, such as education, geography, epidemiology, public health and others (Wu, 1995; Duncan et al., 1996; 1998; Britton, 1990; Carey, 2000; Bryk & Raudenbush, 1992; Paterson & Goldstein, 1992; Carr-Hill et al., 1996; Goldstein, 1995; Leyland & Boddy, 1997; Davis & Gribben, 1995).

Health economic data from multinational trials falls naturally under hierarchical or clustered structure, where the patients are "nested" within the jurisdictions they are treated, e.g. trial sites, hospitals, countries. Most recent health economic literature has extended simple regression approach with the application of hierarchical models to economic data (i.e. costs, effectiveness, incremental net benefit and incremental cost-effectiveness ratio) derived from multinational trials (Rice & Jones, 1997; Manca et al., 2005; Grieve et al., 2005; 2007; Pinto et al., 2005; Willan & Kowgier, 2007; Willan et al., 2005; Manca et al., 2007; Thompson et al., 2010). Hierarchical modelling is appropriate analytical framework to address generalizability. It goes beyond indication of heterogeneity with homogeneity test and allows for estimation of trial-wide and country-level cost-effectiveness parameters and adjustment for patient- and country-level covariates (Drummond et al., 2009).

Assumption made in hierarchical models

By using hierarchical models following assumptions are made:

- Random grouping.

For levels above the lowest level units, e.g. patient-level for clinical or health economic studies, the groups are assumed to be random sample from underlying population. This means that clinical study sites are assumed to be selected randomly from all clinical institutions (DiPrete & Forristal, 1994; Hox & Kreft, 1994).

- Exchangeability.

The important assumption of hierarchical models is the "exchangeability" of the parameters of interest, which are drawn from a common distribution. This assumption implies, that there are no a priori reasons to believe that estimates are relatively high or relatively low in a particular jurisdiction. Obviously, this assumption relates to the above requirements of random grouping.

- Independent blocks.

Hierarchical modelling allows for "clustering" of individual observations within higher-level unit, however the clusters, formed by the subject variables, are assumed to be independent.

- Adequate sample size.

There is no simple rule of thumb for calculation of sufficient sample size. The number of units on both contextual and compositional levels should be tentatively higher if the relationship between different parameters is not strongly pronounced (Goldstein, 1995; Garson, 2009; Spiegelhalter et al., 2002).

Model formulation in the case study

For our case study we adopted the model's formulation proposed by Nixon and Thompson (2005) and Willan and Kowgier (2007) as summarize below.

$E_{ji} \sim Dist(\phi_{E_{ji}}, \sigma_{E_{ji}})$ and $C_{ji} \sim Dist(\phi_{C_{ji}}, \sigma_{C_{ji}})$ where E_{ji} and C_{ji} are the observed effectiveness and costs with respective mean and standard deviation values for patient i in treatment arm j from country $k \in [1:M]$.

Bivariate hierarchical model for country-by-treatment interaction for mean cost and effectiveness is specified respectively:

$$\phi_{E_{ji}} = \mu_{E_j} + \sum_{k=1}^{M} (\gamma_{Ek} + \delta_{Ek} j) z_{jik}$$

$$\phi_{C_{ji}} = \mu_{C_j} + \sum_{k=1}^{M} (\gamma_{Ck} + \delta_{Ck} j) z_{jik} + \beta_j (E_{ji} - \phi_{Eji})$$

where z_{jik} is an indication function for patient i in the treatment arm j from country k. The model assumes that effectiveness may influence costs, through non-zero parameter β_j, however, there is no opposite effect of costs influencing effectiveness. This is a plausible assumption for the fully protocolized clinical study, where expenditures for the trial are fully covered by pharmaceutical company (this assumption may not be valid while analyzing observational data).

For each country there are corresponding regression coefficients $\gamma_{Ek}, \delta_{Ek}, \gamma_{Ck}, \delta_{Ck}$, which assumed to be random-effects with mean zero drawn from bivariate normal distribution. The trial-wide treatment effectiveness is $\Delta_e = \mu_{E1} - \mu_{E0}$ and cost

$\Delta_C = \mu_{C1} - \mu_{C0}$; the country-level treatment and cost effects are $\Delta_e + \delta_{Ek}$ and $\Delta_C + \delta_{Ck}$ respectively.

Covariates x_{ij} are incorporated as following to account for covariate interaction with treatment:

$$\phi_{E_{ji}} = \mu_{E_j} + \sum_{k=1}^{M} (\gamma_{Ek} + \delta_{Ek} j) z_{jik} + (\theta_E + \omega_E j) x_{ji}$$

$$\phi_{C_{ji}} = \mu_{C_j} + \sum_{k=1}^{M} (\gamma_{Ck} + \delta_{Ck} j) z_{jik} + \beta_j (E_{ji} - \phi_{Eji}) + (\theta_C + \omega_C j) x_{ji}$$

Here, the regression coefficients θ and ω are respective direct and interaction effects on costs and effectiveness. The positive interaction effect increases mean incremental cost or effectiveness by ω per unit of covariate. Continues covariates x_{ij} are centered around the trial-wide mean.

Effectiveness is a Bernoulli parameter and is restricted between 0 and 1 as following:

$$\phi_{Eji} = \min(\max(\alpha_{ji}, -\tau_{ji}), 1 - \tau_{ji}) + \tau_{ji}$$

where

$$\alpha_{ji} = \mu_{E0} + \sum_{k=1}^{M} \gamma_{Ek} z_{jik} \text{ and}$$

$$\tau_{ji} = \left(\mu_{E1} - \mu_{E0} + \sum_{k=1}^{M} \delta_{Ek} z_{jik} \right) j$$

with addition of covariate $\tau_{ji} = \left(\mu_{E1} - \mu_{E0} + \sum_{k=1}^{M} \delta_{Ek} z_{jik} \right) j + (\theta_C + \omega_C j) x_{ji}$

Cost data is modelled using gamma distribution with probability density function:

$$f(C_{ji} \mathrm{I} \phi_{Cji}, \rho_{ji}) = \frac{1}{\Gamma(\rho_{ji})} \left(\frac{\rho_{ji}}{\phi_{Cji}} \right)^{\rho_{ji}} C_{ji}^{\rho_{ji}-1} \exp\left(-\frac{\rho_{ji} C_{ji}}{\phi_{Cji}} \right), C_{ji} \geq 0, \rho_{ji} > 0, \phi_{Cji} > 0$$

The gamma distribution allows mean cost to be modeled as linear function of the treatment, covariate and country effects. The shape parameter is unique for each

country, namely: $\rho_{ji} = \sum_{k=1}^{M} z_{jik}\rho_k^*$ (Nixon&Thompson, 2005; Willan&Kowgier, 2007).

.

Variance component model

Model for the cost variable could be presented as following:

$$\phi_{C_{jik}} = \mu_{C_j} + u_{Ck} + e_{Cik}$$

Where $\phi_{C_{jik}}$ is the cost for the ith patient receiving either treatment or comparator in the country k; μ_{C_j} is the intercept, i.e. the average cost from pooling all the observations in the dataset; u_{Ck} is the random term applying to the patients from country k; e_{Cik} is another random term applying to ith patient within the country k.

The model has two random components, namely e_{Cik} and u_{Ck} with zero means and respective variance σ_u^2 and σ_e^2, the variance components, which can be used to calculate intraclass correlation coefficient (ICC) as following:

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

The value of ICC ranges between 0 and 1 and indicates the fraction of total variance that can be attributed to between-country variation (Manca et al., 2005).

Shrinkage estimation

Country-level estimates (slope and intercept) are facilitated through shrinkage estimation. The random term u_{Ck} is a latent variable and is derived in the parameter estimation process. The raw residual of variance component model can be calculated as

$$r_{ijk} = \phi_{C_{ijk}} - \hat{\phi}_{C_{ijk}}$$

where $\phi_{C_{ijk}}$ - observed cost of ith patients from country k and $\hat{\phi}_{C_{ijk}}$ is estimated cost .

The raw residual r_k of the country kth is calculated as the mean of r_{ijk} over all patients in the country. Shrinkage estimation of the country-specific residuals is calculated as following:

$$\hat{u}_{jk} = \frac{\sigma_u^2}{\left[\sigma_u^2 + (\sigma_e^2 / n_k)\right]} \times r_k$$

The multiplier r_k ranges between 0 and 1 so that the estimated residual cannot exceed the raw residual for a specific country. This multiplier is referred to as the "shrinkage factor", which will be noticeably less than one, when σ_e^2 is large relatively to σ_u^2, or when n_k is small or both. In either case we have relatively little information about the particular country, where the outcomes are very variable or there are just few patients, and raw residuals are "shrunk" toward the overall mean. The countries with small sample size and/or unreliable estimates are "borrowing information" from the other countries with less uncertainty in their estimates. Shrinkage estimation can be considered as compromise between fully pooled and fully splitted approach, trial-wide mean and the country-level estimates (Goldstein, 1999; Manca et al., 2005; Willan et al., 2005).

Estimation of hierarchical models

Hierarchical models can be estimated in different ways. Widely used method is maximum likelihood, which rely on assumption of normally distributed random coefficients of the model and requires large dataset. Another commonly used option is estimation via generalized least squares or an iterative version of it, which will provide same results as maximum likelihood, if random effects are normally distributed. Latest development of computational options allowed for full Bayesian analysis with employment of Markov Chain Monte Carlo (MCMC) methods using Gibbs sampling, which are applied to estimate our case study models using the software WinBUGS (Goldstein, 1999; Manca et al., 2005). The main differences between traditional and Bayesian approaches are outlined in the next section.

Bayesian analysis

Traditional "frequentist" estimation with maximum likelihood and Bayesian analysis as implemented in WinBUGS software with MCMC procedure, are distinctive in their underlying basis. In this section, we discuss only selected aspects about the understanding of parameters, distribution and inference. The important difference between Bayesian and frequentist approaches is that Bayesians refer to the model parameters as randomly distributed unknown values with some probability distribution.

In contrast, frequentists understand parameter values as unknown, but fixed without probability. The difference in understanding of the nature of the parameters is translated into distinctive interpretation of probability. Bayesian statistic refers to probability as a "degree of belief", whereas frequentist statistics rely on relative frequency, e.g many repeating observations (e.g. "throwing the dice" many times). The distinctions between understanding of parameters and probability are reflected in different approaches to inference. P-value is essential for frequentist analysis and describes conditional probability to observe more extreme values provided the studied hypothesis is true P(Y $\geq$ c | H0 is true). Bayesian analysis focus on validation of conditional probability based on known evidence and observed data. Conditional probability P(H1 is true | Y) describes probability of the event provided available evidence, knowledge or opinions. Posterior probability adjusts our "degree of belief" into particular event given the old evidence updated with new information.

Formal expression of Bayes' theorem is the following:

$P(\eta \mid Y) \propto P(Y \mid \eta)\, P(\eta)$, where η represents the model parameters, Y the observed data, $P(\eta \mid Y)$ the posterior density function.

The posterior density function contains probability of different value sets for η given the observed data and is basis for Bayesian statistical inference. It is proportional to the product of the likelihood of the data given the model parameter $P(Y \mid \eta)$ and the marginal density for η, $P(\eta)$. Prior density function $P(\eta)$ is informed by available evidence or prior knowledge. As the point estimate for η in WinBUGS, we may take the mode (analogous to the maximum likelihood), the mean or the median of this density. The interval of 2.5th and the 97.5th percentile of the posterior density provides 95% credibility intervals. Credibility intervals are corresponding to "frequentist" confidence intervals in the large datasets (van der Berg et al., 2006; Fryback et al., 2001; Spiegelhalter et al., 1996).

Advantages of Bayesian approach for the case study analysis

One advantage of application Bayesian MCMC for parameter estimates in our case study is that it calculates outcomes in an appropriate form for estimation of cost-effectiveness acceptability curves (Fenwick et al., 2004). Cost-effectiveness acceptability curves provide the probability that the intervention is cost-effective as a function of willingness-to-pay of the decision maker. This data can be estimated from

the posterior distribution of incremental net monetary benefit (Manca et al., 2005). The other advantage of MCMC is it particular application to small samples. Maximum likelihood estimation rests on assumption that studied sample size is sufficiently large. Using MCMC in small samples directly accounts for uncertainty around the values of the random parameters and provides the exact measures of this uncertainty with respective credibility intervals (Goldstein, 1999).

3.7 Summary of Chapter 3

This section provided overview of methods to improve generalizability used in published health economic evaluations and international guidelines, and presented the methodology used in patient-level case study.

Generalizability of health economic evaluations is recognized issue in health economic literature and international guidelines.

Review of health economic guidelines and published health economic studies indicates that the issues is acknowledged by academics, industry and international decision-makers. Understanding of variability and improving generalizability of health economic evidence is supporting efficient resource allocation in research and informed decision-making in policy.

Various clinical and wider non-clinical variability factors are potential sources of between-country heterogeneity.

Number of publications describes patient, disease, health care system and methodological factors, which may contribute to differences in cost-effectiveness results across locations and over the time. Several methods are proposed to deal with variability and to improve generalizability in patient-level studies and decision models.

Decision modelling is complementary analytical "vehicle" to the patient-level studies to improve generalizability, as it facilitates synthesis of data from different sources.

Decision models provide the framework to apply generalizability methods to construct cost-effectiveness results for the long-term therapies, simulate results for the non-trial locations and adjust trial results to reflect routine clinical practice. Models facilitate

calculation of setting-specific cost-effectiveness, also for the cases, where the relative treatment effect is not available from the particular setting. Generalizability of input parameters collected in other locations is subject to uncertainty, which can be appropriately reflected and tested within sensitivity analysis in decision modelling framework.

Patient-level studies should address generalizability at study design, analysis and report phases. Homogeneity test and regression methods are suggested for data analysis.

In the patient-level studies, the generalizability issue is to be considered at the study design stage in order to minimize artificial research bias. Several statistical tools, such as homogeneity tests and regression based methods, were applied in the analysis of the selected multinational studies. Further development of hierarchical regression methods is anticipated avenue for future research on generalizability methods. The appraisal methods and checklists are available to assess relevance of published evidence for decision-making in a specific setting.

Only few studies considered generalizability in the analysis. This topic should be addressed in future research.

The application of the generalizability techniques in the decision modelling and patient-level studies could have different implications for analysis and results across clinical indications and therapies. Available methods were developed and/or applied in particular studies in specific clinical indications. Further research needed to test these methods in other clinical "environment" in the different patient-level studies.

Homogeneity test and hierarchical models are applied in the case study to improve generalizability of the outcomes collected in multinational setting.

Case study is based on secondary analysis of the patient-level data collected in multinational context to compare efficacy of two antibiotic treatments. We selected homogeneity test and hierarchical models to explore between-country variability and potential improvement of generalizability within and beyond study setting.

4 Case Study

In previous sections we reviewed theoretical foundations of health economic methods and practical recommendations and implementation of different approaches to address generalizability issue in international guidelines and health economic literature. Following chapter will demonstrate application of selected generalizability methods in the secondary analysis of the multinational patient-level study. We first provide motivation and objectives of the case study, followed by description of dataset and case study results.

4.1 Case Study Objectives

Purpose of research and case study objectives

The purpose of the thesis is to explore the methods to improve generalizability of health economic evaluations and can be specified in the terms of two following research objectives:

1) To identify methods to explore heterogeneity and to improve generalizability applied in published health economic evaluations conducted in multinational settings.

2) To apply selected methods to the patient-level data in a case study to investigate between-country variability of cost-effectiveness parameters, heterogeneity factors explaining between-country heterogeneity and to calculate country-level and trial-wide cost-effectiveness estimates.

We addressed the first objective in the literature review, described in the previous section. Homogeneity test and hierarchical models are recommended by recent health economic literature as tools to explore variability and to improve generalizability. The need to gain more inside about using these methods for health economic analysis in different indications, treatment patterns and settings has been acknowledged. To address the second research objective, we apply selected methods in the case study, based on secondary analysis of a multinational study to investigate efficacy and safety of sequential intravenous/oral moxifloxacin (the treatment) in comparison to sequential

intravenous/oral co-amoxiclav (the comparator) with or without clarithromycin in the treatment of patients with community-acquired pneumonia.

Analytical approach for the case study could be essentially described by the following steps: (1) assessment of between-country variability, (2) quantification of major heterogeneity factors, (3) estimation of country-level and trial-wide parameters and (4) based on heterogeneity factors and country-level estimates, assessment of feasibility to generalize results to non-study settings. In detail, the case study research objectives are specified in Table 4.

Table 4: Research Objectives of the Case Study

Nr	Objectives	Methods
1	To provide crude mean and incremental costs and effectiveness estimates based on trial-wide and country-level patient data analysis.	Descriptive statistics
2	To provide crude analysis of patient-level and country-level heterogeneity factors, which are likely to have effect on costs and effectiveness estimates.	Descriptive statistics
3	To assess qualitative country-by-treatment interaction in incremental costs, effectiveness, resource use and NMB.	Qualitative homogeneity test
4	To assess quantitative country-by-treatment interaction in incremental costs, effectiveness, resource use and NMB.	Quantitative homogeneity test
5	To determine degree of within-country clustering.	Variance component model ICC
6	To calculate trial-wide and country-level estimates of the incremental cost and effectiveness using patient-level data while accounting for within country clustering.	Hierarchical models for cost-effectiveness

Nr	Objectives	Methods
7	To calculate trial-wide and country-level estimates of the incremental cost and effectiveness using patient-level data while accounting for within country clustering and controlling for patient-level and country-level covariates.	Hierarchical models for cost-effectiveness with covariates
8	To explore and assess model fit after country- and patient-level covariates adjustment.	Hierarchical models with and without covariate
9	To assess covariates effects on mean and incremental costs and effectiveness estimates and possibility to generalize results of the study to non-study settings on hand of explored covariates (external generalizability).	Hierarchical models with covariates
10	To calculate cost-effectiveness acceptability curves based on estimates from models without and with covariates.	Hierarchical models with and without covariate
11	To calculate trial-wide and country-level estimates of the incremental LOS using patient-level data and accounting for within country clustering.	Hierarchical models for LOS
12	To calculate trial-wide and country-level estimates of LOS using patient-level data and accounting for within country clustering while controlling for patient-level and country-level covariates.	Hierarchical models for LOS with covariates
13	To explore and assess model fit after country- and patient-level covariates adjustment.	Hierarchical models for LOS with and without covariates

Nr	Objectives	Methods
14	To assess covariates effects on mean and incremental LOS estimates and possibility to generalize results of the study to non-study settings on hand of explored covariates (external generalizability).	Hierarchical models for LOS with covariates

The case study research is of explorative nature and is not deemed for primarily cost-effectiveness analysis. Thus, no a priori hypothesis are formulated, if not otherwise specified.

4.2 Dataset

4.2.1 Data Sources and Management

Case study is based on a secondary analysis of a multinational, prospective, randomised, open study to investigate the efficacy and safety of sequential intravenous/oral moxifloxacin (the treatment) in comparison to sequential intravenous/oral co-amoxiclav (the comparator) with or without clarithromycin in the treatment of patients with community-acquired pneumonia requiring initial parenteral treatment (BSP study report, 2000). The study and all clinical work were conducted in accordance with Good Clinical Practice for Trials on Medicinal Products in the European Community (EMEA, 2002). Local and Central Ethic Votum were obtained before the start of the study as well as patient's informed consent to participate in this clinical trial. The results of primarily analysis are published elsewhere (Finch et al., 2002; Drummond et al., 2003).

Patient-level clinical database, report and unit costs for UK, Spain, Germany and France for secondary analysis for application of generalizability methods were received from Bayer Schering Pharma. Data management, collection of unit costs for Switzerland, country-level covariates and complete analysis was performed by the author. Resource use valuation and costing was performed in Eclipse JAVA application.

Calculation of total costs and resource utilization was done by the author and may deviate from results of primary analysis.

Descriptive statistics and calculation of ICC have been performed with Stata software package version 9 (StataCorp, 2005). Descriptive analysis (frequencies, means, and standard deviations) was done to examine the demographics and baseline characteristics of the subjects. Explorative bivariate inferential statistics was performed in form of t-test for continuous variables and chi2 test for binary or categorical variables. The a priori level of significance was 0.05 for all statistical tests. All tests were two-tailed unless otherwise specified. Calculation of ICC was performed by fitting the variance component model using loneway procedure in Stata without covariates.

Estimation of hierarchical models was facilitated by Markov Chain Monte Carlo using WinBUGS software. WinBUGS code was adapted from the model by Willan and Kowgier (2007) and is given in the appendix. If not otherwise specified, the posterior mean estimates and 95% credibility intervals were obtained from MCMC estimation running 2 chains for minimum 50,000 iterations following a burn-in period of a minimum 20,000 iterations. WinBUG's Gelman-Rubin diagnostics was used to validate convergence (Spiegelhalter et al., 2003).

Incremental cost and effectiveness were estimated using 3 hierarchical models: model 1 (HM1) without covariates, model 2 (HM2) with complete set of country- and patient-level covariates, model 3 (HM3) with the patient-level covariates only. Analogue to incremental cost and effectiveness models, we estimated incremental LOS using 3 hierarchical models: model 4 (HM4) without covariates, model 5 (HM5) with complete set of country- and patient-level covariates, model 6 (HM6) with the patient-level covariates only. Crude or "naïve" estimates are calculated using country-specific data sub-samples.

Bivariate correlation for continuous covariates was done to check for extreme multicollinearity. Results are provided in appendix. Notably, some degree of collinearity could be tolerated, since the continuous covariates are centered around trial-wide means. However, no evidence for strong linear correlation between the covariates could be found.

4.2.2 Indication

Community-acquired pneumonia is an infectious disease acquired in social and community places other than clinical or long-term care institutions. The morbidity and mortality associated with CAP is high. In Germany, approximately 350,000 to 500,000 cases of CAP are reported each year (Ewig et al., 2002), while over 700,000 patients per year are affected in France (Mark, 2008). Study conducted in US estimated incidence as of 5.6 million cases per year, of those ca. 20 % required inpatient care (Niederman et al., 1998; 2001). In UK, about 50,000 CAP patients annually required hospitalization (British Thoracic Society, 1993). Mortality for inpatient population is 2-21% and increases to over 50% for the severe CAP patients (Marrie, 1998; Bartlett & Mundy, 1995).

In approximately 30 to 60% of CAP patients diagnostic tests fail to identify the targeted pathogen, therefore the empiric approach to the CAP therapy is needed. Guidelines developed by the British and American Thoracic Societies in 1993 recommend that empiric therapy for CAP to provide coverage against Streptococcus pneumoniae, the most common cause of CAP (British Thoracic Society, 1993; American Thoracic Society, 1993).

The economic impact of CAP on the health care systems is high because of associated effects, such as excess of hospitalization for certain underlying conditions, heart failure and long-standing deterioration of health. Burden of illness study conducted in UK estimated direct health care costs at approximately 441 GBP million per year in 1992-1993. Inpatient care contributed to major proportion of the total annual cost. The mean costs of outpatient clinical case was ca. 100 GBP, whereas inpatient management required 1700-5100 GBP (Guest & Morris, 1997).

A recent study from Spain analyzed information from the national surveillance hospital data during two-years period. Study concluded that CAP accounted for 53,000 hospitalizations with cost to the Spanish health care system of approximately 140 million USD per year (Monge et al., 2001).

In the US, the annual CAP costs were estimated around 8.4 billion USD. Approximately half of the total costs was paid for hospital services. The mean duration

of hospitalization varied between 5.8 days for those under 65 years of age and 7.8 days for older patients (Niederman et al., 1998).

Antibiotics account for a small portion of the total costs of CAP treatment. Given the differential for hospital versus outpatient treatment in most countries, treatments, that reduce the length of hospitalization, may result in significant cost-savings for the health care system. This concept has been confirmed in several economic studies in which CAP treatments, that resulted in shorter lengths of stay, created cost-savings without affecting patient outcomes (Palmer et al., 2000; Dresser et al., 2001; Coughlin et al., 2003).

4.2.3 Clinical Study Design

The clinical study was a multinational, multicentral, randomized, open label, phase III study designed to examine the safety and efficacy of sequential intravenous (IV)/oral (PO) moxifloxacin (the treatment) compared to a standard regimen of sequential IV/PO co-amoxiclav with or without clarithromycin (IV or PO) (the comparator) in the treatment of CAP in adult patients requiring initial parenteral therapy. Collection and assessment of resource use data was a secondary objective of this clinical study. The study started in February 1999 and completed in May 2000, enrolling a total of 628 patients. Patients were screened and randomized at 65 centers from 10 countries: Germany (19% of patients), Greece (16%), Israel (13%), South Africa (12%), and France (10%), followed by the United Kingdom (9%), Switzerland (8%), Spain (7%), Belgium (4%), and Russia (2%).

Clinical characteristics of the patient population enrolled into the study included the following:

- Patients aged 18 years or above.
- Patients requiring initial parenteral treatment whose clinical condition suggested that they would require at least 3 days of intravenous therapy.
- Patients who were willing and able to provide written informed consent.
- A diagnosis of CAP was to be confirmed on an out-patient basis or within 48 hours after hospital admission.

- Signs and symptoms of pneumonia (patients were to have evidence of categories, A,B and C):

A. Fever (core temperature > 38.5C and/or oral temperature > 38C) and/or leucocytosis (WBC > 10000/mcl) and/or left shift (>=15% band forms).

B. One or more of the following characteristics: productive cough, purulent sputum, dyspnoea or tachypnoea (>20 breaths/minute), rigors/chills, pleuritic chest pain, auscultation findings consistent with pulmonary consolidation.

C. Radiological evidence of an infiltrate consistent with pneumonia.

Disease severity was stratified as following:

Stratum 1: The status of mild to moderate community-acquired pneumonia was assigned to patients not eligible for stratum 2.

Stratum 2: Severe community-acquired pneumonia: presence of at least one of the following conditions:

- Respiratory rate >30 breaths/minute
- Hypoxaemia with PO2 <8 kPa (60 mmHg)
- Requirement for mechanical ventilation
- Chest radiograph showing bilateral involvement or involvement of multiple lobes
- Diastolic blood pressure <=60 mmHg
- Requirement for vasopressors for more than 4 hours

Patients received either treatment or comparator intravenously for a minimum of 3 days. At any time after this, following assessment by the investigator, the patient could be switched to oral therapy. Treatment was to last for a minimum of 7 days, and for a maximum of 14 days (Finch et al., 2002; BSP study report, 2000).

4.2.4 Patient Population

622 patients (99.0%) of the 628 patients enrolled were correctly randomised and received at least one dose of study medication, hence they were eligible for the intention-to-treat population (ITT). 301 patients were in the treatment group and 321 patients were in the comparator group. 6 patients were excluded from the ITT analysis: 1 patient was excluded due to a randomisation error, 4 patients were excluded since they received no study medication and 1 patient was excluded due to previous enrolment in the study.

Following numbers of patients were observed in each race class for the ITT population:

"Caucasian"	507 patients	(81.5%)
"Black"	48 patients	(7.7%)
"Asian"	3 patients	(0.5%)
"Other"	1 patient	(0.2%)
"Not documented "	63 patients	(10.1%)

In France local regulations prohibit the documentation of race. Similar profile is being observed within the two treatment groups. The baseline comparability of the treatment groups is presented in Table 5.

Table 5: Baseline Parameters of Treatment Groups

Variable, categories		Mean ± standard deviation or frequencies			p-value
		Treatment	Comparator	Total	
Sex	Male	193 (64.1%)	207 (64.5%)	400 (64.3%)	0.92
	Female	108 (35.9%)	114 (35.5%)	222 (35.7%)	
Disease severity	Mild/Mod	143 (47.5%)	158 (49.2%)	301 (48.4%)	0.67
	Severe	158 (52.5%)	163 (50.8%)	321 (51.6%)	

Age (years)		(55.2 ± 20.6)	(55.9 ± 19.6)	(55.6 ± 20.1)	0.65
Weight (kg)		(72.2 ± 15.6)	(72.4 ± 17.1)	(72.3 ± 16.4)	0.85
Temperature (°C)		(38.7 ± 0.9)	(38.8 ± 1.0)	(38.8 ± 1.0)	0.46
Respiration rate (bpm)		(24.4 ± 7.0)	(25 ± 7.1)	(24.7 ± 7.1)	0.22
Hospitalisation	No	148 (49.2%)	156 (48.6%)	304 (48.9%)	0.89
pre-therapy	Yes	153 (50.8%)	165 (51.4%)	318 (51.1%)	
Pre-existing broncho-pulmonary disease	No	217 (72.1%)	229 (71.3%)	446 (71.7%)	0.84
	Yes	84 (27.9%)	92 (28.7%)	176 (28.3%)	
Smoking	Never	123 (40.9%)	122 (38.0%)	245 (39.4%)	
History	Passive	1 (0.3%)	3 (0.9%)	4 (0.6%)	
	Previous	91 (30.2%)	92 (28.7%)	183 (29.4%)	
	Smoker	86 (28.6%)	104 (32.4%)	190 (30.5%)	
Alcohol consumption	Abstinent	156 (51.8%)	159 (49.5%)	315 (50.6%)	0.41
	Light	110 (36.5%)	116 (36.1%)	226 (36.3%)	
	Moderate	29 (9.6%)	31 (9.7%)	60 (9.6%)	
	Heavy	6 (2.0%)	14 (4.4%)	20 (3.2%)	
	No data		1 (0.3%)	1 (0.2%)	
Cough	None	20 (6.6%)	26 (8.1%)	46 (7.4%)	0.89
	Mild	82 (27.2%)	90 (28.0%)	172 (27.7%)	
	Moderate	151 (50.2%)	149 (46.4%)	300 (48.2%)	
	Severe	48 (15.9%)	56 (17.4%)	104 (16.7%)	
	Present	1 (0.3%)	1 (0.3%)	2 (0.3%)	

Dyspnoea	None	57 (18.9%)	53 (16.5%)	110 (17.7%)	0.7
	Mild	87 (28.9%)	87 (27.1%)	174 (28.0%)	
	Moderate	95 (31.6%)	112 (34.9%)	207 (33.3%)	
	Severe	62 (20.6%)	69 (21.5%)	131 (21.1%)	
	Present		1 (0.3%)	1 (0.2%)	

Of the 622 patients valid for the intention-to-treat population 400 (64.3%) were male (193 and 207 in the treatment and comparator treatment groups respectively) and 222 (35.7%) were female (108 and 114 in the treatment and comparator treatment groups respectively). Of the 622 intention-to-treat patients 301 patients (48.4%) (143 and 158 in the treatment and comparator treatment groups respectively) had mild to moderate CAP and 321 (51.6%) of the ITT patients had severe CAP (158 and 163 in the treatment and comparator treatment groups respectively). Three hundred and eighteen of the 622 ITT patients (51.1%) were hospitalized pre-therapy (153 (50.8%) and 165 (51.4%) in the treatment and comparator treatment groups respectively). There was no statistically significant difference between treatment groups in relation to the incidences of hospitalization, p=0.8859. The majority of patients had coexisting medical histories, 258 (85.7%) and 282 (87.9%) in the treatment and comparator treatment groups respectively. The most frequently observed conditions related to the endocrine, nutritional, metabolic and immunity system (69 (22.9%) and 87 (27.1%) in the treatment and comparator treatment groups respectively), circulatory system (135 (44.9% and 132 (41.1%) in the treatment and comparator groups respectively) and respiratory system (124 (41.2%) and 129 (40.2%) in the treatment and comparator treatment groups respectively). 84 (27.9%) and 92 (28.7%) of the treatment and comparator treatment groups respectively had coexisting broncho-pulmonary disease giving a total of 176 (28.3%). 451 patients were febrile at the start of the study (temperature >= 38.5 C) (215 (71.4%) and 236 (73.5%) in the treatment and comparator groups respectively). Overall, the two treatment groups were well comparable with regard to demographic and baseline characteristics (findings verified with BSP study report, 2000).

4.2.5 Estimation of Effectiveness Outcome

Patients enrolled in the study were followed until 21 to 28 days after the last study drug intake. The maximum time period to the test of cure (TOC) visit of 21 days (i.e. 14 days maximum treatment followed by 7 days follow-up) was selected for the economic evaluation time frame since it corresponded to the primary clinical efficacy measure (i.e. clinical assessment at TOC visit).

Efficacy endpoints were defined as following:

- Clinical cure: Resolution of clinical symptomathology related to the infection without requirement of further antibacterial therapy.
- Clinical failure: Death due to CAP, failure or partial response to the study drug treatment, requiring a modification or extension of antibacterial therapy.
- Indeterminate: Patients in whom a clinical assessment was not possible to determine (e.g. due to early withdrawal from the study because of adverse events, protocol violation, withdrawn consent, etc.).

The overall treatment effectiveness measure was the proportion of patients cured based on the primary efficacy parameter: the clinical response to study drug at the TOC visit (i.e. 5 to 7 days post-treatment) (Finch et al., 2002). The most conservative approach was used to estimate clinical effectiveness, considering all patients with the outcome of indeterminate or missing as clinical failures.

4.2.6 Estimation of Resource Use and Costs

The economic analysis was conducted from the perspective of the health care providers: the mandatory German Sickness Funds (Gesetzliche Krankenkassen Versicherung (GKV)), the French public health insurance sector, the Spanish National Health System, the UK National Health Service (NHS) and Swiss Statutory Health Insurance. The timeframe of economic evaluation are years 1999 - 2001. It was assumed that all patients in the study would have been eligible for coverage under each of these public health insurance plans. Health economic analysis from societal

perspective was not performed due to unavailability of data for indirect costs. In following we outline organization of health care in respective countries in 1999-2001 and summarize costing approach for individual resource use.

Description of health care systems in UK, Spain, Germany, France and Switzerland

In Germany, health care is funded by membership of sickness funds. The health care system requires that all working people have health insurance provided by the mandatory sickness funds (GKV), with membership fees divided equally between employer and employee. The GKV covers a wide range of health care services including hospitalizations, inpatient procedures, rehabilitation, and medications, with respective patient co-payments associated with each health care service.

The French health care system is primarily funded by national sickness insurance funds, which are part of the social security system. The insurance funds are financed by compulsory payroll contributions of employers and employees. National insurance provides for a wide range of outpatient and hospital services, pharmaceutical therapies and other essential care needs. Approximately 75% of health expenditures are accounted for by the insurance funds, with the remainder covered by the central government, patient co-payments, and Mutual Insurance Funds, which provide private supplemental insurance coverage.

The Spanish National Health System is financed out of general taxation. Coverage is almost universal for citizens of Spain and includes extensive list of services and care to all citizens regardless of personal wealth. Benefits include general medical care at doctors' offices and patients' homes, outpatient and inpatient care including all medical and surgical specialties in acute care, prescription medications with patient co-payments of up to 40% for those received out of hospital, and complementary benefits such as prostheses, orthopedic products, and wheelchairs. Individuals may chose to have membership in an alternative, employment-linked insurance program, and also have the option of purchasing additional private insurance.

The NHS in the UK is financed out of general taxation, employers and employees via the National Insurance Scheme. All UK citizens are provided coverage for inpatient care, outpatient care, and medications. In addition, is it possible to buy private health

care insurance which is fully paid by the insured or which can be provided (as a benefit) by employers to employees, in which case the employer bears all costs.

Swiss health care is funded by a system of public-private statutory health insurance, where country's 26 cantons (regions) and more than 3000 communes (municipalities) have substantial power with regard to financing and management of the health care system. Individual cantons act as the main financers for around 80% of hospitals within the public sector. Health care spending is financed primarily by the Statutory Health Insurance system (SHI) and supplementary private insurance. SHI system generates around 40% of total health expenditure. Around 35% of the population purchases voluntary supplementary health care insurance in addition to the statutory insurance system for dental, travel, and other items not covered via the statutory system's basic basket of services. Procedures related to treatment of acute conditions, such as CAP, would be covered by statutory service to a large extend (IHS Global Insight, 2008; Drummond et al., 2003; BSP study report, 2000; WHO European Centre for Health Policy, 2008).

Estimation of CAP-related resource use and cost

Resources were valued using 1999-2001 unit costs and reimbursement rules from Germany, France, Spain, UK and Switzerland to determine CAP-related cost to each country's health care system. Unit costs for Germany, France, UK and Switzerland were collected in German Deutsche Marks (DM), French Francs (FF), British Pounds and Swiss Francs (SFr) and total costs for each resource component were converted into EUR using the conversion rates of 1 DM = 0.5113 EUR, 1 FF = 0.1524 EUR, 1GBP = 1.5674 EUR and 1SFr = 0.657 EUR (Bank of Canada, 2008). Unit costs for Spain were collected in EUR. Exchange rates were used, since they more accurately reflect what a domestic buyer has to pay for a product or service within the country. The valuation of costs with purchasing power parity for West European Countries would not have major impact on estimates due to very small variation in coefficients between the countries (Musgrove & Fox–Rushby, 2006; OECD, 2009).

The CAP-related cost included study medication, hospitalizations, concomitant and follow-up CAP-related medications, radiological procedures, and therapeutic adjuncts over the 21-day evaluation time frame. Protocol-related costs associated with study visits were not included into the total cost estimation. From clinical perspective, both

evaluation timeframe as well as study procedures are considered representative for general medical practice for CAP treatment with antibiotics. Economic analysis was integrated part of the research objective and protocol-driven utilization and sufficient observation timeframe were considered during the study design phase (Drummond et al., 2003; BSP study report, 2000). Following section summarize major aspects of costing approach, detailed unit costs and data sources are provided in attachment.

Estimation of the study medication costs

Utilization of study medication was determined for each patient based on the compliance record within the case report form where the day, time, and route of each dose were recorded. The number of days of study medication use was determined considering all days within the 21-day evaluation time frame upon which at least one dose of study medication was received. The cost per patient was determined by multiplying the actual number of doses received by the cost per dose. Although the cost of medications received in hospital are likely to differ from those received out of hospital due to contract negotiations between hospitals and suppliers, this information is not publicly available and would be highly variable between the hospitals. Therefore, the public price of study medications was assumed for all doses received both in hospital and following hospital discharge. Only doses received out of hospital were subjected to local reimbursement rules.

For comparator medications in Germany, a weighted cost was assumed for the oral forms of co-amoxiclav and clarithromycin based on the current market share of brand and generic name products. For co-amoxiclav IV, the price of branded product was assumed. For the study medication received out of hospital, a patient co-payment of 4.60 EUR per package was subtracted to obtain the cost to the German health care system.

For the French analysis, the costs of branded products were used for the IV and PO forms of co-amoxiclav and clarithromycin. A reimbursement rate of 65% was applied to all doses received out of hospital to represent the cost to the public health insurance funds.

For the comparator medications in Spain, a weighted cost was assumed based on the current market share of brand and generic name products. For all medications received out of hospital, patients pay 40% of the cost with the exception of specific groups for

which there are no out-of-pocket payments. These groups include the retired (age 65 and over), the handicapped, invalids, and people who have suffered occupational accidents. Although it was not possible to identify the proportion of study patients who were handicapped, invalid, or had suffered occupational accidents, the proportion of study patients under the age of 65 years was determined and a weighted reimbursement rate of 76% was calculated. Hence, for all doses of study medications received out of hospital, a reimbursement rate of 76% was applied to obtain the cost to the Spanish health care system.

For the UK analysis for all comparator medications, a weighted cost was assumed based on the current market share of brand and generic name products with the prices obtained from official sources: the Drug Tariff and the British National Formulary. These two sources also list prices for generic medications. Prescription pharmaceuticals are not subject to value added tax. For medications dispensed in the community, 100% reimbursement was assumed. In practice, this assumption is not strictly true because non-exempt patients pay a co-payment of 9.72 EUR per prescription item. However, it was not practicable to identify the proportion of non-exempt patients in the study and normal UK practice in economic evaluations is to assume 100% reimbursement. In hospital, the true acquisition prices of medications are not known as listed prices are subject to discounts negotiated either with the wholesaler or the manufacturer. It was therefore assumed that the published basic NHS prices prevailed in hospital.

For Swiss analysis, costs of the medication to the health insurance were reduced by 10% co-pay. Hospital prices of the drugs are most likely to be different from the public prices. Suppliers have a special interest that their drugs are prescribed in hospital since this is generally the first location where drugs are prescribed. Due to confidentiality of the price negotiations, hospital prices for medications are generally not publicly available. Thus, public price of the pharmaceuticals is assumed for in-patient care (HIS Global Insight, 2008; BSP study report, 2000; WHO European Centre for Health Policy, 2008).

Estimation of hospitalization duration and costs

The number of days of hospitalization by ward type within the 21-day time frame was determined for each patient based on the number of days between the admission and discharge dates recorded in the case report form (CRF). Any readmissions within the

evaluation time frame were also included in the total. The cost of hospitalization was determined for each patient by multiplying a daily cost of hospitalization by the number of days spent in each ward. Separate per diems were obtained for stays in an intensive care unit (ICU) and in a general ward. All wards that were not identified as ICU in the CRF were assumed to have the same cost as the general ward since costs were not available for all ward types listed. Hospitalization costs included resources such as treatments received in hospital, medical procedures, laboratory tests, physician services, nursing care, board and lodging.

The costs obtained in Germany represented the average cost for all admissions, and a patient co-payment of 8.69 EUR per day was subtracted from the per diem for the first 14 days (PKV, 1999/2000).

The costs obtained in France were based on weighted average per diem costs for appropriate diagnostic related groups in the public and private not-for-profit sectors, and a direct payment made by the patient to the hospital of 8.38 EUR per day was not included in the daily cost of hospitalization to the French health care sector (PMSI, 2000).

The costs obtained in Spain represented the average cost for CAP and related conditions from a variety of sources including literature estimates and hospitals (SOIKOS, 2001).

For the UK, fees for hospitalization were determined using the NHS Reference Costs 2001 database, which consists of costs obtained from the NHS hospitals. The per diem values used in the study represented the average cost for bronchopneumonia patients treated in NHS trust hospital (NHS Reference costs, 2001/2002).

In Switzerland, basic general ward and ICU per diem cost were taken based on Taxordnung for Kanton St. Gallen (Taxordnung, 2002).

Since the hospitalization costs obtained from each country were fully allocated and included the cost of medications, in our analysis we removed the proportion of the total per diem represented by antibiotics. We included patient-specific study medication costs, which vary between treatment groups and may have impact on cost-effectiveness results.

Estimation of total hospital costs and cross-country comparison is challenging due to different remuneration systems (DRG, per diem, capitation reimbursement contracts) employed in each country. In addition, there is variation due to local contracts between hospitals and e.g. municipal administrations. Thus, the costs of the comparable clinical cases may widely vary within and across countries. In the case study we calculated total costs assuming per diem remuneration in all countries assuming same unit costs for all hospital in the same country. This estimation may not truly reflect the actual costs per clinical case. However and since the assumption was consistent across all hospitals and countries, this imprecision should not have major implication for studying generalizability methods.

Costs for concomitant medications

Out-patient use of CAP-related concomitant and post-therapy antimicrobials and other CAP-related concomitant medications was determined for each patient within the 21-day evaluation time frame based on the start and stop dates recorded in the CRF. Only medications relevant to the direct treatment of CAP, as judged by a clinician, were included in the analysis. This included antimicrobials prescribed for pneumonia as well as other medications prescribed for the following indications: cough and congestion, dyspnea, fever, hypotension, pain, tachypnea, tachycardia, short ventricular tachycardia, bradycardia, pulmonary insufficiency, respiratory distress or insufficiency, cardiogenic shock (in case of septic shock), cardiac failure, respiratory acidosis, renal insufficiency related to CAP, electrolyte substitution, hyperkalemia, hypokalemia, dehydration, low magnesium, low phosphate, sepsis, and septic shock (BSP study report, 2000). Public prices for all CAP-related medications reimbursed by the German, French, Spanish, Swiss and UK health care systems were obtained from the following sources. For Germany, the source was the Rote Liste® (2001); for France: Vidal-Semp (2001); for Spain: Vademecum International (2001) and Portalfarma (2001); for Switzerland: Documed (2001) and for the UK: British National Formulary (2001) and NHS Drug Tariff (2001). Unit prices for the daily dosages consumed were multiplied by the durations of use within the time frame to obtain the cost for each medication. Appropriate reimbursement rates were applied to each medication cost to determine the cost from each health care perspective. Medications received in hospital were not valued separately as the hospitalization costs obtained were fully allocated and included medication costs.

Costs for out-patient radiological procedures

The number of out-patient, non-protocol driven radiological procedures conducted for the treatment of CAP was determined for each patient within the evaluation time frame. Relevant procedures reported by patients included chest X-ray and chest CT scan (Finch et al., 2002). The cost of each procedure was obtained from each country.

For Germany, the source was Gebührenordnung für Ärzte (GOA, 2000); for France: National Fee Schedule (NGAP, 2001) and l'Union des Caisses Nationales de Sécurité Sociale (TIPS, 2001); for Spain: Base de datos de costes sanitarios (SOIKOS, 2001); for Switzerland: Tarmed (2002) and for the UK: NHS Drug Tariff (2001).

After applying the appropriate reimbursement rules, the cost was multiplied by the number of procedures completed to determine the cost of this component to each health care system. Inpatient radiological procedures were not valued separately since the cost of these services was included in the average daily hospitalization cost.

Costs for out-patient therapeutic adjuncts

The number of days on which therapeutic adjuncts were received on an out-patient basis for the treatment of CAP was determined for each patient within the evaluation time frame. Procedures relevant to the analysis included aerosol breathing therapy, chest physiotherapy, inhalation therapy, nasal c-pap, and oxygen received via face mask or nasal cannula. For adjuncts received by a physiotherapist, it was necessary to estimate the number of treatments that would be received per week since only the start and stop dates for the entire treatment period were recorded on the CRFs. It was estimated by a clinician that, on average, patients would receive five 20- to 30- minute sessions per week. The cost of each adjunct (including professional time and supplies, where appropriate and available) was obtained from each country.

For Germany, the source was Physiotherapy offices in Northrhein-Westphalia (2001); for France: National Fee Schedule (NGAP, 2001) and Tarif Interministériel des Prestations Sanitaires (TIPS, 2001); for Spain: Base de datos de costes sanitarios (SOIKOS, 2001); for Switzerland: Physioswiss (2001); and for the UK: NHS Drug Tariff (2001) and Personal Social Services Research Unit (2001).

After applying the appropriate reimbursement rules, the cost was multiplied by the number of days of adjuncts received to determine the cost of this component to each

health care system. Inpatient therapeutic adjuncts were not valued separately since the cost of these services was included in the average daily hospitalization cost. Further details on unit costs estimates are listed in appendix.

4.2.7 Advantages of the Dataset for Studying Generalizability Methods

This dataset offers several advantages for studying between-country variability and application of generalizability methods.

- Heterogeneous multinational setting.

Clinical and economic evidence was collected in standardized way as directed in common clinical protocol across 10 countries, which represent various settings with different clinical and socio-economic environment.

- Acute condition with internationally defined diagnostic procedure.

Community acquired pneumonia is an acute condition which may be considered as to limited extend exposed and influenced by the factors unobserved in clinical trial setting. The analysis of chronic conditions may be more challenging as the patients may get different diagnosis and treatment during the disease course, what may not be captured and studied within the clinical study timeframe.

- Hospitalization is major resource use.

In-hospital stay is the main cost-driving resource use for all health care settings. Inconsistency in costing methodology may introduce additional “technical” variability to the economic parameters. In fact, collection of unit costs for all 10 countries is resource intensive and difficult due to unavailability of unit costs data in some countries. Thus it is feasible to apply generalizability methods directly to resource use parameters such as length of hospitalization.

4.3 Results

This section provides results of the case study analysis per objective as defined in section 4.1. The results are presented as following. We first provide the summary of crude estimates for effectiveness, costs, LOS and covariates, followed by results of between-country qualitative and quantitative heterogeneity test. We then present results of hierarchical modelling for costs and effectiveness for 5 countries and LOS for the complete dataset of 10 countries. Finally, we present effects of country and patient-level covariates and discuss potential for external generalizability.

4.3.1 Crude Country-Level Estimates

Objective 1: Estimation of crude mean and incremental costs and effectiveness based on trial-wide and country-level data analysis.

Table 6, Table 7 and Table 8 present crude or "naïve" cost, effectiveness and LOS estimates calculated using country-level and trial-wide sub-groups. Cost estimates were calculated for 5 countries with available country-specific unit costs. Major resource utilization – hospitalization – and effectiveness data are calculated for all participating countries. Histograms for costs and LOS are provided to visually check distributional properties and skewness of the outcomes.

Recruitment of patients in different countries was unbalanced resulting in the smallest patient sample in Russia (12 patients) and the largest in Germany (118 patients) (Table 7). Small country-level samples do not allow for robust resource use and cost-effectiveness analysis in all participating countries using the patient data from these respective countries only.

Brief inspection of country-level data for incremental costs, net monetary benefit, effectiveness and LOS in Table 6, Table 7 and Table 8 indicate potential between-country heterogeneity. New treatment is cost saving in Germany, UK, France and Switzerland, but not in Spain. The magnitude of cost saving is strongly pronounced in France (-1191.84 EUR) and UK (-979.16 EUR), to a smaller extent in Switzerland (-219.90 EUR), whereas the data for Spain suggests the comparator is cost saving over the treatment (80.40 EUR). As shown on histograms (Figure 9), cost data are skewed to the right. Similar data shape can be observed for data distribution on treatment- and

country- levels as shown on the diagrams in appendix. Not-normal distributions was confirmed by significant Shapiro-Wilk test ($p< 0.001$). This fact is relevant for the regression analysis. Nevertheless, arithmetic mean and t-test are robust to violation of the normality assumption and could be considered informative for the country-level analysis.

Similarly to the costs, there are differences across the countries in mean and incremental effectiveness (Table 7). Mean country-level effectiveness varies from 0.58 in Belgium to 0.92 in Greece in treatment arm and from 0.62 in Switzerland to 0.86 in Spain in comparator arm. Treatment is more efficacious in Germany, Spain, UK, France, Greece, Israel and Russia, but has no pronounced difference or comparator is more efficacious in South Africa, Belgium and Switzerland.

Both mean and incremental LOS (Table 8) show potential differences between the countries with the longest mean LOS in Russia (12.60 days) and shortest in South Africa (4.03 days) in treatment arm and longest in Germany (12.68 days) and shortest in South Africa (4.20 days) in comparator arm respectively. Similar to the cost, LOS data are also skewed to the right as shown on histograms (Figure 10). Histograms on data distribution per treatment arm are provided in appendix.

Overall, country-level analysis indicates potential between-country differences in incremental effectiveness, costs and LOS, which could be attributed to either patient-level (compositional) characteristics or underlying country-specific (contextual) factors.

Table 6: Summary Data for Costs and Incremental Net Monetary Benefit

Country	N Treat	N Comp	N total	C Treat	SD Treat	C Comp	SD Comp	Diff Treat-Comp	95% CI		Ttest	INMB (1000)
DE	56	62	118	2476.25	1182.54	3033.07	1249.56	-556.82	-1001.65	-111.99	0.02	673.75
ES	22	21	43	2562.89	2036.24	2482.49	2344.68	80.40	-1270.21	1431.01	0.91	-73.91
UK	29	30	59	1683.74	903.65	2662.90	2251.43	-979.16	-1879.37	-78.95	0.03	1139.16
FR	30	34	64	5133.37	4520.42	6325.22	5814.00	-1191.84	-3820.03	1436.34	0.37	1260.84
CH	23	26	49	3088.41	1502.12	3308.31	2330.80	-219.90	-1363.60	923.80	0.70	213.21
5 CNTR	160	173	333	2930.73	2551.08	3590.42	3350.40	-659.69	-1305.51	-13.87	0.05	743.87

Note: Costs are calculated in EUR, N denotes number of patients, C - cost of treatment, INMB – incremental net monetary benefit

Table 7: Summary Data for Effectiveness: Proportion of Cured Patients in Treatment and Comparator Arms

Country	N Treat	N Comp	N total	E Treat	SD	E Comp	SD	Total	SD	Diff	95% CI		P chi2
DE	56	62	118	0.88	0.33	0.76	0.43	0.81	0.39	0.12	-0.020	0.254	0.10
ES	22	21	43	0.86	0.35	0.86	0.36	0.86	0.35	0.01	-0.201	0.213	0.95
UK	29	30	59	0.79	0.41	0.63	0.49	0.71	0.46	0.16	-0.067	0.386	0.18
FR	30	34	64	0.83	0.38	0.76	0.43	0.80	0.41	0.07	-0.127	0.263	0.50
CH	23	26	49	0.61	0.50	0.62	0.50	0.61	0.50	-0.01	-0.280	0.267	0.96
BE	12	11	23	0.58	0.52	0.64	0.51	0.61	0.50	-0.05	-0.451	0.345	0.79
GR	50	50	100	0.92	0.27	0.84	0.37	0.88	0.33	0.08	-0.046	0.206	0.220
IL	39	40	79	0.77	0.43	0.73	0.45	0.75	0.44	0.04	-0.147	0.236	0.650
RU	5	7	12	0.80	0.45	0.71	0.49	0.75	0.45	0.09	-0.399	0.570	0.740
ZA	35	40	75	0.80	0.41	0.80	0.41	0.80	0.40	0.00	-0.182	0.182	1.000
5 CNTR	160	173	333	0.81	0.39	0.73	0.45	0.77	0.42	0.08	-0.006	0.170	0.07
10 CNTR	301	321	622	0.81	0.39	0.75	0.43	0.78	0.41	0.06	-0.001	0.128	0.06

Table 8: Summary Data for Hospitalization (LOS)

CNTR	N Treat	N Comp	N total	LOS Treat	SD Treat	LOS Comp	SD Comp	Total	SD	Diff	95% CI		P ttest
DE	56	62	118	10.73	5.55	12.68	4.88	11.75	5.28	-1.94	-3.85	0.04	0.05
ES	22	21	43	7.41	5.07	7.38	5.48	7.39	5.21	0.03	-3.22	3.28	0.97
UK	29	30	59	6.10	3.63	8.63	4.82	7.39	4.43	-2.53	-4.76	0.30	0.03
FR	30	34	64	8.87	6.64	10.65	6.71	9.81	6.68	-1.78	-5.12	1.56	0.29
CH	23	26	49	9.61	4.46	9.35	5.41	9.46	4.94	0.26	-2.61	3.13	0.85
BE	12	11	23	7.67	4.64	9.36	4.03	8.48	4.64	1.70	-2.35	5.75	0.39
GR	50	50	100	6.84	4.00	7.50	4.07	7.17	4.03	-0.66	-2.26	0.94	0.41
IL	39	40	79	5.95	4.68	5.00	2.60	5.47	3.78	0.95	0.74	2.64	0.27
RU	5	7	12	12.60	6.42	8.29	5.94	10.08	6.26	4.31	3.69	12.32	0.26
ZA	35	40	75	4.03	2.58	4.20	2.50	4.12	2.53	-0.17	-1.34	1.00	0.77
5 CNTR	160	173	333	8.93	5.49	10.43	5.70	9.71	5.64	-1.51	-2.72	-0.30	0.02
10 CNTR	301	321	622	7.63	5.19	8.43	5.37	8.05	5.29	-0.81	-1.64	0.03	0.06

Note: LOS are calculated in days

Figure 9: Cost Histogram for Treatment and Comparator Arms Combined

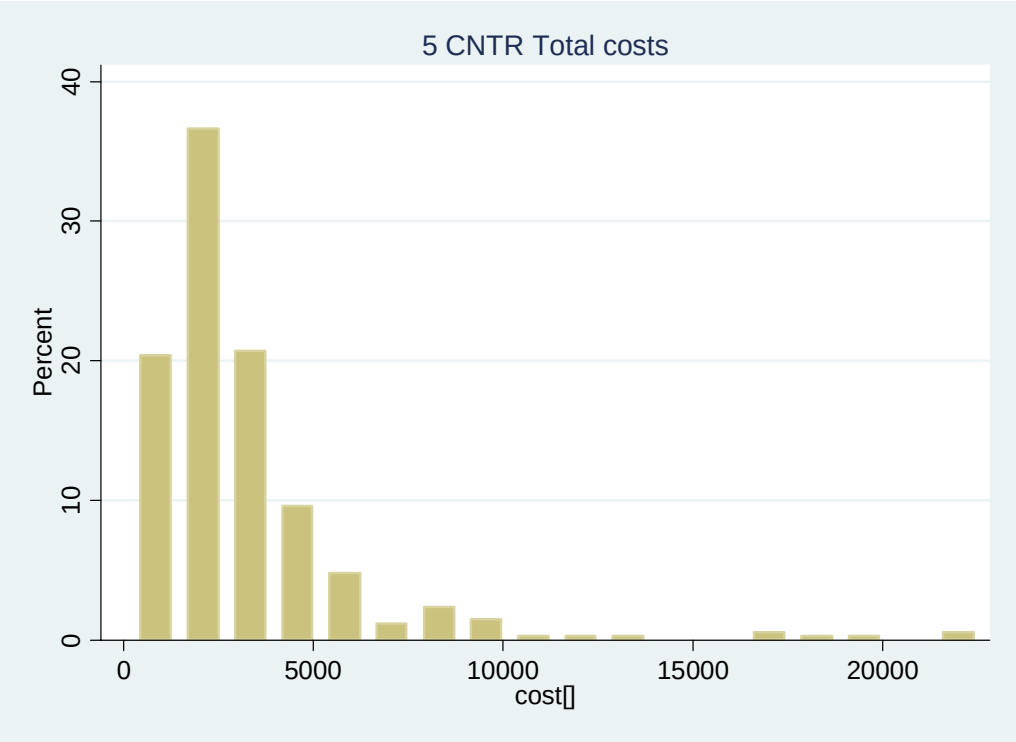
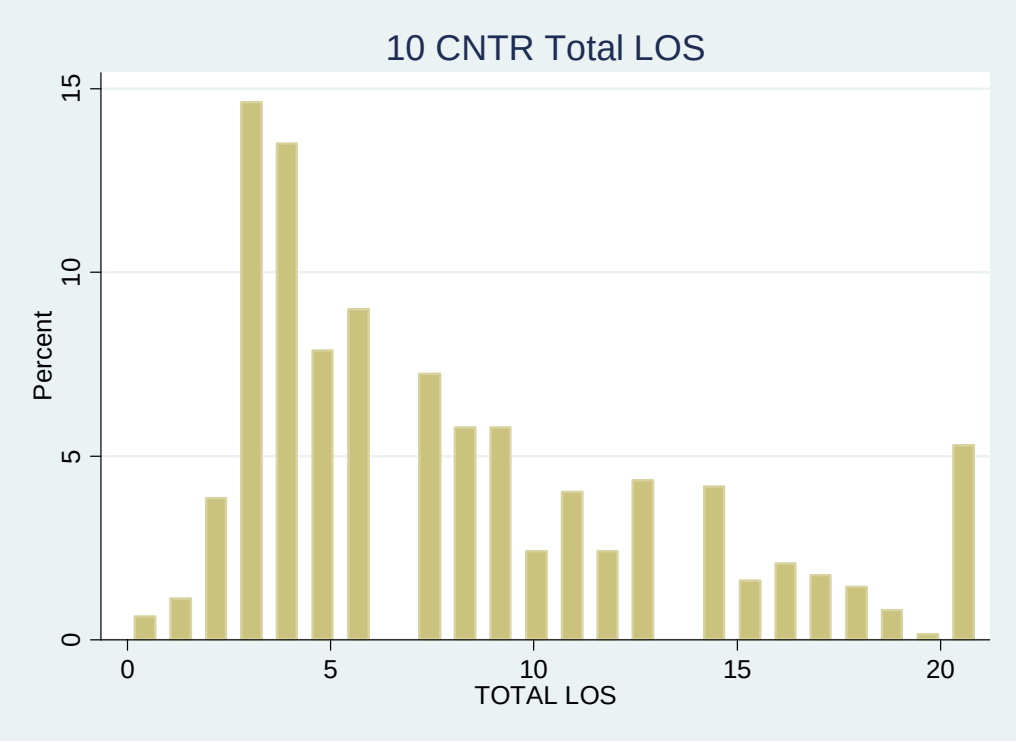


Figure 10: LOS Histogram for Treatment and Comparator Arms Combined



Objective 2: Crude analysis of patient-level and country-level covariates (heterogeneity factors) which are likely to have effect on mean and incremental costs, resource use and effectiveness outcomes

Patient-level and country-level covariates were selected based on the conclusions of the literature review of generalizability methods (section 3) and selected health economic studies in CAP indication (Fine et al., 1990; 1993; 1997; 2000; Battleman et al., 2002; McCormick et al., 1999; Woodhead et al., 2004; Ruiz et al., 1999; Janssens, 2005; Lave et al., 1999).

Following patient-level covariates were hypothesized to have effect on resource use and costs:

- demographic characteristic: patients age
- clinical characteristic: severity of CAP condition at hospital admission
- health-related behavior: mean number of cigarettes smoked per day.

Respectively, following country-level covariates were thought to have effect on resource use and costs:

- characteristics related to national health-care resources: total health expenditure as % of gross domestic product (GDP) and number of hospital beds per 100 000 inhabitants
- characteristic related to clinical presentation of disease on national level: bacteria resistance to antibiotic treatment.

Summary statistics on patient- and country-level covariates is presented in Table 9, Table 10, Table 11 and Table 12. Patient-level covariates were collected during the study course. Health expenditures as % of GDP and number of hospital beds per 100 000 for individual countries were extracted from WHO HFA database (WHO, 2009). Measurement and quantification of antibiotic resistance on country-level appeared to be a complex issue (Low, 2004). On hand of EARRS (2006) report we have dichotomized countries with high and low level of antibiotic resistance (level of covariate 1 and 0 respectively).

Results show variability in country-level characteristics. The highest total health expenditures as proportion of GDP are observed in Switzerland (10.5%) followed by Germany (10.3%) and France (9.7%) in contrast to Russia (2.8%) with lowest total health expenditure. Even greater difference can be seen in number of hospital beds per 100 000 population with highest hospital capacity in Russia (1085 beds/100 000 inhabitants) and lowest in Spain (377.6 beds/100 000 inhabitants). Overall, examination of selected country-level characteristics suggests substantial differences between the settings as measured by selected parameters. Thus, further analysis should deem to explore and quantify association between the country-level characteristics, mean and incremental costs and resource use for treatment of CAP in the particular settings.

Table 9: Summary Statistics for Patient- and Country-Level Covariates

Variable	Description	Source	DE	ES	UK	FR	CH	BE	GR	IL	RU	ZA
Patient-level covariates												
CX2	Severity of CAP condition, 1=severe CAP, 0=mild CAP	data base	0.43	0.70	0.37	0.48	0.73	0.39	0.40	0.49	0.33	0.79
CX4	Mean patient's age	data base	58.6	60.3	53.4	60.8	67.4	48.3	57.3	56.1	35.6	40.6
CY3	Mean number of cig/day	data base	5.8	4.3	3.3	5.8	7.4	8.4	7.6	6.0	12.6	3.1
Country-level covariates												
CX3	Antibiotics resistance, 1=high resistant, 0=low	EARSS, 2006	0	1	0	1	0	1	1	1	1	0
CY1	Total health expenditure as % of gross domestic product (GDP), 1999	WHO, 2009	10.3	7.3	7.1	9.7	10.5	8.6	7.8	8.2	2.8	7.2 (*)
CY2	Number of hospital beds per 100000 population in 1999	WHO, 2009	919.6	377.6	416.4	828.7	660.8	559.2	472.4	612.0	1085.0	480.0(*)

Note: *WHO data not is not available. Information extracted from published sources: United Nations (2008), van Walbeek (2002)

Descriptive statistics for patient-level covariates - severity of CAP, age and smoking - is provided in Table 10, Table 11 and Table 12. As expected in randomized design study, there is no substantial difference in proportion of patients with severe CAP between treatment (0.53) and comparator arm (0.51) in the pooled set over 10 countries. However, the proportion of patients with severe CAP varies from 0.20 in Russia to 0.86 in South Africa and 0.30 in UK to 0.73 in South Africa in treatment and comparator arms respectively. The relative difference between the arms varies from 0.15 in UK to -0.30 in Belgium. There is no indication for statistically significant difference within the countries between treatment and comparator arms. However, the absolute and insignificant between-country differences may impact mean and incremental resource use and cost. This association is investigated in multivariate analysis.

Table 10: Descriptive Statistics on Proportion of Patients with Severe CAP at Admission

Country	N Pat	CAP Treat	SD	CAP Comp	SD	Total	SD	Diff	95% CI		Ch2 p value
DE	118	0.45	0.50	0.42	0.50	0.43	0.47	0.03	-0.15	0.21	0.77
ES	43	0.68	0.48	0.71	0.46	0.70	0.47	-0.03	-0.31	0.24	0.82
UK	59	0.45	0.51	0.30	0.47	0.37	0.49	0.15	-0.1	0.39	0.24
FR	64	0.47	0.51	0.50	0.51	0.48	0.50	-0.03	-0.28	0.21	0.79
CH	49	0.78	0.42	0.69	0.47	0.73	0.45	0.09	-0.15	0.34	0.48
BE	23	0.25	0.45	0.55	0.52	0.39	0.50	-0.30	-0.68	0.09	0.15
GR	100	0.40	0.50	0.40	0.50	0.40	0.49	0.00	-0.19	0.19	1.00
IL	79	0.49	0.51	0.50	0.51	0.49	0.50	-0.01	-0.23	0.21	0.91
RU	12	0.20	0.45	0.43	0.54	0.33	0.49	-0.23	-0.74	0.27	0.41
ZA	75	0.86	0.36	0.73	0.45	0.79	0.41	0.13	-0.05	0.31	0.16
5CNTR	333	0.53	0.50	0.49	0.50	0.51	0.50	0.04	-0.07	0.15	0.47
10CNTR	622	0.53	0.50	0.51	0.50	0.52	0.50	0.02	-0.06	0.1	0.67

Similar to severity of CAP condition, and owing to randomized study design, there is no substantial difference in mean patients' age between treatment (55.21 SD 20.51) and comparator arm (55.94 SD 19.59) in the trial-wide set. However, mean age varies from 33.00 years to 68.61 years and 37.43 years and 66.23 years in Russia and Switzerland in treatment and comparator arms respectively. There is no indication for significant difference in mean patient age between treatment and comparator arms per country and trial-wide. However, ANOVA analysis (normality assumption confirmed by Shapiro-Wilk test) indicates significant $p<0.001$ between-country difference in mean age for both 5 countries and 10 countries sets.

Table 11: Descriptive Statistics for Patients' Age at Admission

Country	N Pat	Treat	SD	Comp	SD	Total	SD	Diff	95% CI		p value
DE	118	56.73	17.69	60.19	17.44	58.55	17.57	3.46	-2.95	9.87	0.28
ES	43	59.91	17.53	60.76	15.05	60.33	16.18	0.85	-9.23	10.93	0.87
UK	59	54.14	22.60	52.60	20.99	53.36	21.62	-1.54	-12.90	9.83	0.79
FR	64	55.43	22.46	65.50	15.95	60.78	19.78	10.07	0.42	19.72	0.04
CH	49	68.61	16.36	66.23	14.74	67.35	15.40	-2.38	-11.31	6.56	0.60
BE	23	50.75	16.31	45.64	21.88	48.30	18.91	-5.11	-21.75	11.53	0.53
GR	100	56.52	23.27	58.02	19.38	57.27	21.32	1.50	-7.00	10.00	0.73
IL	79	56.72	19.74	55.50	19.34	56.10	19.42	-1.22	-9.97	7.54	0.78
RU	12	33.00	10.27	37.43	14.94	35.58	12.60	4.43	-12.89	21.75	0.58
ZA	75	42.83	17.20	38.45	16.20	40.49	16.71	-4.38	-12.07	3.31	0.26
5CNTR	333	58.16	19.73	60.90	17.59	59.58	18.67	2.73	-1.29	6.76	0.18
10CNTR	622	55.21	20.51	55.94	19.59	55.57	20.03	0.74	-2.42	3.90	0.65

There is insignificant difference in mean number of cigarettes smoked per day between treatment (5.36) and comparator arm (6.23). Mean number cigarettes varies from 2.89 in UK to 12.00 in Russia and 2.90 in Spain and 13.00 in Russia in treatment and comparator arms respectively.

In early simulations with hierarchical models we considered to use smoking consumption data as country-level covariates. Number of cigarettes smoked per person per day may not be directly considered as contextual effect. However, indirectly, mean average consumption per person can reflect country-level contextual factors, such as social acceptance of smoking, effect of political/financial incentives against smoking, etc. Table 13 presents descriptive summary of cigarettes consumption calculated using study dataset and data extracted from WHO-HFA database. There is only small deviation between mean number of cigarettes smoked per person/day in study dataset (5.81) and WHO data (5.14). However, country-level statistics shows substantial differences between country-level study data set and WHO data for Germany, Spain, France, Belgium and Russia. This example indicates potential problematic of collecting country-level covariates outside the study and applying these to small patient samples selectively recruited into the study. In fact, mean parameter values reported on higher country-level may not be valid for the population included in the study.

Table 12: Descriptive Statistics for Mean Daily Cigarette Consume

Country	Nr Pat	Treat	SD	Comp	SD	Total	SD	Diff	95% CI		p value
DE	118	5.45	11.05	6.27	11.40	5.88	11.20	0.83	-3.28	4.93	0.69
ES	43	5.50	10.02	2.90	9.55	4.33	9.76	-2.60	-8.62	3.44	0.39
UK	59	2.89	6.98	3.67	7.54	3.29	7.21	0.77	-3.02	4.56	0.69
FR	64	6.50	10.76	5.15	7.93	5.78	9.31	-1.35	-6.04	3.33	0.57
CH	49	7.39	17.64	7.38	11.55	7.39	14.56	-0.01	-8.48	8.47	0.99
BE	23	7.67	14.49	9.18	15.85	8.39	14.83	1.52	-11.64	14.67	0.81
GR	100	5.54	10.26	9.70	14.86	7.62	12.88	4.16	-0.91	9.23	0.11
IL	79	5.18	11.42	6.75	11.47	5.97	11.40	1.57	-3.56	6.70	0.54
RU	12	12.00	10.96	13.00	14.60	12.58	12.65	1.00	-16.30	18.30	0.90
ZA	75	3.06	5.70	3.18	6.36	3.12	6.02	0.12	-2.68	2.91	0.93
5CNTR	333	5.36	10.00	5.47	11.41	5.42	10.68	-0.11	-2.42	2.20	0.93
10CNTR	622	5.36	10.83	6.23	11.19	5.81	11.01	0.88	-0.86	2.61	0.32

Table 13: Comparison of Smoking Consumption Based on Study Set and WHO Data

Country	Nr Pat	Study mean cig consume patient/day	SD	WHO cig/person/year 2000	WHO person/day	Diff Study-WHO
DE	118	5.88	11.20	1553.15	4.26	1.62
ES	43	4.33	9.76	2464.44	6.75	-2.42
UK	59	3.29	7.21	1123.39	3.08	0.21
FR	64	5.78	9.31	1303.29	3.57	2.21
CH	49	7.39	14.56	2336.36	6.40	0.99
BE	23	8.39	14.83	1532.50	4.20	4.19
GR	100	7.62	12.88	2953.80	8.09	-0.47
IL	79	5.97	11.40	2161.00	5.92	0.05
RU	12	12.58	12.65	2411.14	6.61	5.97
ZA	75	3.12	6.02	933.00	2.56	0.56
5CNTR	333	5.42	10.68	1756.13	4.81	0.61
10CNTR	622	5.81	11.01	1877.21	5.14	0.67

Note: * WHO consumption for Israel is based on 2003 data, all other countries on 2000

Source: Study database and WHO, 2009

4.3.2 Test of Between-Country Heterogeneity

Objective 3: Assessment of qualitative country-by-treatment interaction in incremental costs, effectiveness, resource use and INMB.

The results of qualitative homogeneity including test statistic and test power are summarized in Table 14. The results of qualitative homogeneity test suggest absence of country-by-treatment interactions in incremental effectiveness, length of stay, costs and INMB. That indicates, essentially, that the effect's direction appears to be homogeneous in all countries, i.e. new treatment is more efficacious, saves hospitalization time and

costs. However, the test power for incremental effectiveness, cost, INMB and LOS (5 countries) is low. Particularly low test power is seen for incremental costs and INMB (0.41 and 0.51 respectively), where the variability of estimates is high, and thus, the heterogeneity may remain undetected.

Table 14: Results of Qualitative Homogeneity Test

Incremental effect	Test Statistic		Critical Value	Power
	Q+	Q-		
Effectiveness 5 CNTR	5.3	0	4.96	0.58
Costs 5 CNTR	0.01	10.4	4.96	0.41
NMB 5 CNTR	11.62	0.01	4.96	0.51
LOS 5 CNTR	1.53	2.84	4.96	0.68
LOS 10 CNTR	5.94	6.52	10.99	0.96

Objective 4: Assessment of quantitative country-by-treatment interaction in incremental costs, effectiveness, resource use and NMB.

Test for quantitative interactions, similar to the results of the test for qualitative interaction, suggests absence of between-country heterogeneity for incremental effectiveness, costs and INMB with p-values of 0.80, 0.70 and 0.67 respectively (see Table 15). However, the results for incremental length of hospitalization indicate significant quantitative between-country heterogeneity, suggesting significant difference in the magnitude to which the treatment reduces length of hospitalization in different countries.

Table 15: Results of Quantitative Homogeneity Test

Country	Incremental Efficacy	Incremental Costs	INMB (1000)	Incremental LOS 5CNTR	Incremental LOS 10CNTR
DE	0.12	-556.82	673.75	-1.95	-1.95
ES	0.01	80.40	-73.91	0.03	0.03
UK	0.16	-979.16	1351.61	-2.53	-2.53
FR	0.07	-1191.84	1047.79	-1.78	-1.78
CH	-0.01	-219.90	213.21	0.26	0.26
Test stat H	1.67	2.20	2.37	0.21	0.60
P value	0.80	0.70	0.67	0.00	0.00

Overall, the tests indicated homogeneity of effect's direction and magnitude. However, due to modest test power for qualitative interactions and significant between-country heterogeneity stated by the test for quantitative interactions in incremental LOS, potential between-country heterogeneity can not be completely ruled out. Hence, it remains unclear, to which extend could trial-wide cost, effectiveness and resource use estimates be considered informative for the decision making in individual countries.

4.3.3 Country-level Cost-Effectiveness Estimates While Accounting for Data Hierarchy and Heterogeneity

Objective 5: Determination of degree of within-country clustering

Intraclass correlation coefficient was calculated to explore degree of clustering within the countries in LOS and costs estimates. Variance component models were fitted in STATA using "loneway" procedure. The results are summarized in Table 16.

Table 16: Estimation of ICC for Economic Parameters Cost and LOS

Setting	Nr patients	Parameter	ICC	95% CI	
5 CNTR	333	Cost	0.203	0.000	0.452
5 CNTR	333	LOS	0.111	0.000	0.272
10 CNTR	622	LOS	0.228	0.038	0.418

Intraclass correlation is measuring the strength of nesting (or clustering), e.g. the extent to which observations depend on higher-level variable, which is country in our case study. Higher value of ICC suggests the use of hierarchical modelling instead of conventional regression techniques. Regression without accounting for hierarchical structure may provide misleading outcomes of significance test due to lack of independence of the error terms (Garson, 2009). The presented results suggest the evidence for clustering within the countries for the parameters costs and LOS. The results in Table 16 are indicative. Summary of the components of variance is appropriate for the normally distributed data, but is problematic for skewed distributions where within-group measure is based on skewed data, but not the between-group measure. Thus, direct interpretation of ICC for given dataset is not feasible.

Objective 6: Calculation of trial-wide and country-level estimates for the incremental cost and effectiveness using patient-level data while accounting for within country clustering.

The results of incremental effectiveness and costs for 5 countries estimated with hierarchical model 1 without covariates are reported in Table 17. Trial – wide estimates for incremental effectiveness is 0.081 CR -0.007 – 0.169 and incremental costs -555.50 CR -945.00 - -160.10 suggesting that new treatment is more efficacious and less costly. This result is consistent with country-level estimates for all countries. The country-level incremental cost and effectiveness estimates are shrunk towards trial-wide results. For each country k there are four regression coefficients estimated with the model: $\gamma_{Ek}, \delta_{Ek}, \gamma_{Ck}, \delta_{Ck}$ where δ_{Ek}, δ_{Ck} are random effects for incremental costs and effectiveness drawn from bivariate normal distribution. The random effects are re-centered so that weighted sum is zero and are interpreted as country-level effect of trial-

wide main and interaction effects; extend of shrinkage depends on estimated precision within each country and true heterogeneity between the countries.

Table 17: Results of Hierarchical Model 1 without Covariates

Node	Setting	Mean	SD	MC error	2.50%	Median	97.50%
Incremental costs	Trial-wide	-555.50	199.10	1.31	-945.00	-556.80	-160.10
	Germany	-564.00	212.00	1.16	-983.50	-563.40	-147.60
	Spain	-328.70	432.10	4.59	-1094.00	-367.10	622.40
	UK	-826.60	303.90	2.99	-1477.00	-807.10	-279.50
	France	-547.90	542.50	5.82	-1650.00	-558.90	608.90
	Switzerland	-417.50	398.10	3.29	-1159.00	-441.50	444.70
Incremental effectiveness	Trial-wide	0.081	0.045	0.000	-0.007	0.081	0.169
	Germany	0.095	0.052	0.000	-0.005	0.093	0.200
	Spain	0.083	0.058	0.000	-0.034	0.083	0.197
	UK	0.079	0.057	0.000	-0.039	0.079	0.189
	France	0.081	0.055	0.000	-0.032	0.081	0.188
	Switzerland	0.049	0.072	0.001	-0.126	0.058	0.166

Note: Incremental costs are calculated in EUR, incremental effectiveness is calculated as difference in proportion of cured patients.

Objective 7: Calculation of trial-wide and country-level estimates of the incremental cost and effectiveness using patient-level data while accounting for within country clustering and incorporating patient-level and country-level covariates.

The results of incremental effectiveness and costs for 5 countries estimated by hierarchical model 2 with complete set of country-level and patient-level covariates and model 3 with patient-level covariates only are reported in Table 18 and Table 19.

Trial –wide estimates for incremental effectiveness as calculated with model 2 are 0.073 CR -0.015 – 0.161 and incremental costs -475.10 CR -869.00 - -81.84 suggesting that new treatment is more efficacious and less costly. These results are consistent with incremental effectiveness and costs estimates in individual countries.

The trial-wide results of model 3, 0.073 CR -0.015 – 0.160 for incremental effectiveness and -472.30 CR -839.00 - -100.50 for incremental costs, do not substantially depart from the results of model 2 and are consistent across all countries.

Inclusion of patient-level covariate "age" produced small differences on the effectiveness estimates in models 2 and 3 comparing to model 1 (data not shown here). The preliminary simulations showed that covariates other than "age" showed negligible effect on incremental effectiveness and were therefore not included into final effectiveness equations in model 1-3.

Table 18: Results of Hierarchical Model 2 with Patient-Level and Country-Level Covariates

Node	Setting	Mean	SD	MC error	0.025	Median	0.975
Incremental costs	Trial-wide	-475.10	199.40	2.16	-869.00	-474.90	-81.84
	Germany	-443.50	427.30	16.38	-1292.00	-458.20	498.30
	Spain	-186.60	674.80	23.41	-1452.00	-251.40	1335.00
	UK	-755.10	586.30	22.32	-2129.00	-694.70	314.80
	France	-510.20	634.30	12.47	-1868.00	-503.50	805.90
	Switzerland	-421.40	395.20	5.47	-1203.00	-430.30	397.00
Incremental effectiveness	Trial-wide	0.073	0.045	0.000	-0.015	0.073	0.161
	Germany	0.085	0.051	0.000	-0.013	0.085	0.188
	Spain	0.073	0.057	0.000	-0.043	0.074	0.183
	UK	0.070	0.056	0.000	-0.043	0.071	0.177
	France	0.071	0.055	0.000	-0.040	0.072	0.176
	Switzerland	0.050	0.065	0.001	-0.101	0.056	0.160

Table 19: Results of Hierarchical Model 3 with Patient-Level Covariates

Node	Setting	Mean	SD	MC error	0.025	Median	0.975
Incremental costs	Trial-wide	-472.30	187.90	1.53	-839.00	-473.70	-100.50
	Germany	-481.20	198.20	1.41	-874.40	-480.10	-92.07
	Spain	-232.70	405.30	6.00	-939.70	-270.80	672.00
	UK	-663.10	296.10	3.96	-1293.00	-646.10	-124.20
	France	-481.10	500.70	7.05	-1529.00	-481.20	563.20
	Switzerland	-420.00	350.50	3.76	-1100.00	-431.00	318.50
Incremental effectiveness	Trial-wide	0.073	0.045	0.000	-0.015	0.073	0.160
	Germany	0.085	0.051	0.000	-0.014	0.084	0.187
	Spain	0.073	0.057	0.000	-0.042	0.073	0.183
	UK	0.070	0.056	0.000	-0.043	0.071	0.178
	France	0.071	0.055	0.000	-0.040	0.072	0.177
	Switzerland	0.049	0.066	0.001	-0.105	0.055	0.160

Objective 8: Overall comparison and assessment of DIC across the models with and without covariates adjustment

Table 21 provides overview of the results calculated by models 1 – 3 and descriptive statistics. Figure 11 provides graphical presentation of incremental costs obtained with hierarchical models and descriptive statistics.

Comparing results obtained with descriptive statistic and hierarchical modelling, we observe the largest between-country variation in incremental costs and effectiveness in descriptive statistics estimates. Country-level estimates in hierarchical models are shrunk toward trial-wide results, what represents a “compromise” between aggregated

(trial-wide) or disaggregated (country-level) analysis. The disaggregated analysis (descriptive statistics) may provide more extreme results due to the fact that some countries (e.g. Spain) have fewer patients. Modelling of cost with gamma distribution allows for better data representation and accounting for the small amount of patients producing very high costs. This explains why incremental costs for France are smaller (in absolute amount) if calculated with hierarchical models comparing to descriptive statistics.

The deviance information criteria (DIC) was calculated to indicate and compare data fit across different models (Spiegelhalter et al., 2002). Model with lower DIC is generally regarded as improvement of data fit whereas difference in 5-10 points is considered as substantial improvement (The BUGS Project, 2010). DIC for model 1 is 6216.82, model 2 and 3 maintained better data fit confirmed by a smaller DIC of 6188.36. Comparing the results of model 2 and model 3, we state, that model 3 provides gain in precision of the mean incremental costs estimates as showed by narrower credibility intervals while maintaining the same DIC as the model 2.

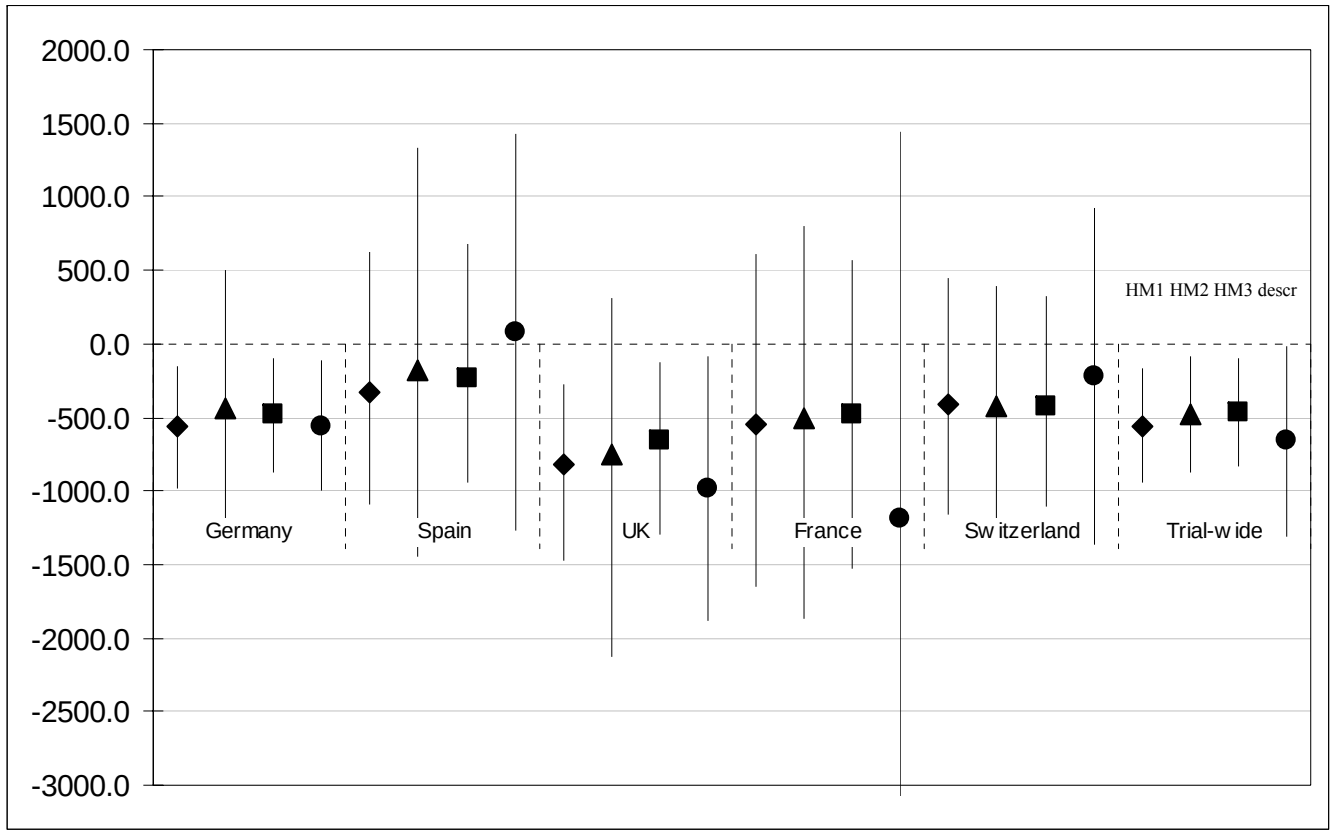
Table 20: Deviance Information Criteria (DIC) for Hierarchical Models 1-3

Model	HM 1 no covariates	HM 2 all covariates	HM 3 pat-level covariates
DIC Costs	5858.01	5830.20	5830.20
DIC Efficacy	358.82	358.16	358.16
DIC total	6216.82	6188.36	6188.36

Table 21: Results of Incremental Cost and Effectiveness Estimated with Hierarchical Models 1-3 and Descriptive Statistics

		HM1: 5 CNTR			HM2: 5 CNTR ALL COV			HM3: 5 CNTR PAT COV			DESCR STAT		
		Mean	CR 2.5	CR 97.5	Mean	CR 2.5	CR 97.5	Mean	CR 2.5	CR 97.5	Mean	CR 2.5	CR 97.5
Incremental cost	Germany	-564.00	-983.50	-147.60	-443.50	-1292.00	498.30	-481.20	-874.40	-92.07	-556.82	-1001.65	-111.99
	Spain	-328.70	-1094.00	622.40	-186.60	-1452.00	1335.00	-232.70	-939.70	672.00	80.40	-1270.21	1431.01
	UK	-826.60	-1477.00	-279.50	-755.10	-2129.00	314.80	-663.10	-1293.00	-124.20	-979.16	-1879.37	-78.95
	France	-547.90	-1650.00	608.90	-510.20	-1868.00	805.90	-481.10	-1529.00	563.20	-1191.84	-3820.03	1436.34
	Switzerland	-417.50	-1159.00	444.70	-421.40	-1203.00	397.00	-420.00	-1100.00	318.50	-219.90	-1363.60	923.80
	Trial-wide	-555.50	-945.00	-160.10	-475.10	-869.00	-81.84	-472.30	-839.00	-100.50	-659.69	-1305.51	-13.87
Incremental effectiveness	Germany	0.095	-0.005	0.200	0.085	-0.013	0.188	0.085	-0.014	0.187	0.117	-0.020	0.254
	Spain	0.083	-0.034	0.197	0.073	-0.043	0.183	0.073	-0.042	0.183	0.006	-0.201	0.213
	UK	0.079	-0.039	0.189	0.070	-0.043	0.177	0.070	-0.043	0.178	0.160	-0.067	0.386
	France	0.081	-0.032	0.188	0.071	-0.040	0.176	0.071	-0.040	0.177	0.069	-0.127	0.263
	Switzerland	0.049	-0.126	0.166	0.050	-0.101	0.160	0.049	-0.105	0.160	-0.007	-0.280	0.267
	Trial-wide	0.081	-0.007	0.169	0.073	-0.015	0.161	0.073	-0.015	0.160	0.080	-0.006	0.170
DIC total		6216.82			6188.36			6188.36			NA		

Figure 11: Incremental Cost Estimated with Hierarchical Models and Descriptive Statistics



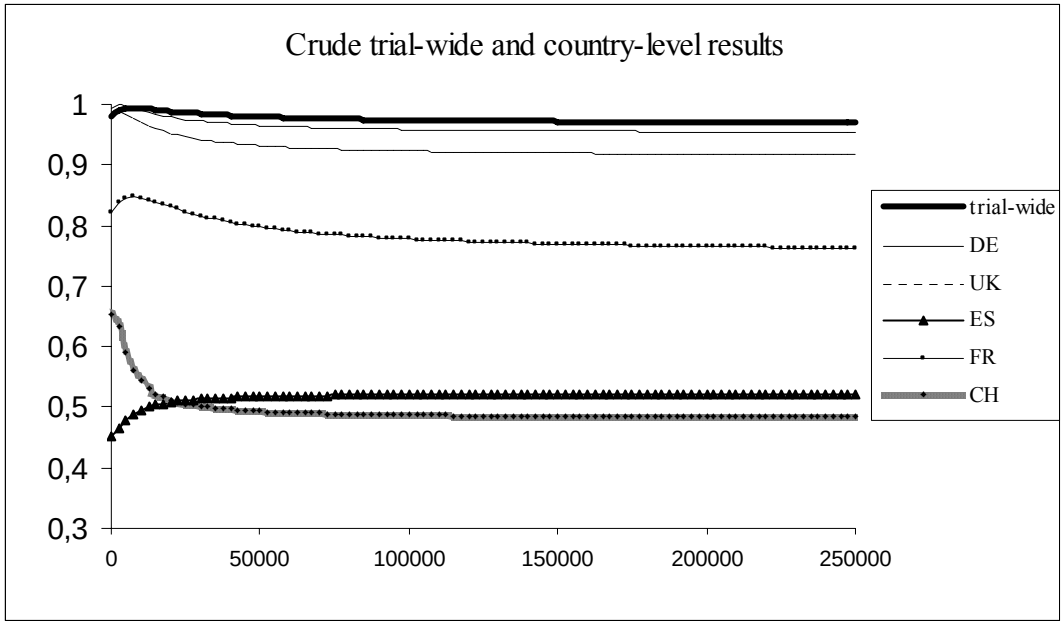
Note: Incremental costs (Y axis) are calculated in EUR

4.3.4 Cost-Effectiveness Acceptability Curves

Objective 9: Construction of cost-effectiveness acceptability curves based on estimates from models without and with covariates.

The INMB is defined as $INMB = \Delta C\lambda - \Delta E$, where lambda is willingness to pay of a payor for a unit of effectiveness. Figure 12, Figure 13, Figure 14 and Figure 15 display cost-effectiveness acceptability curves, i.e. the probability that new therapy is cost effective as a function of lambda (Fenwick et al., 2004). Formally, there will be no difference in decision using the different estimates of hierarchical models 1 to 3, since the probability that new treatment is cost-effective remain above 0.5 at any willingness-to pay value in all countries. However, the ability to calculate country-level estimates by borrowing information from the other countries is a clear advantage and allow for flexible decision making using country-specific results and willingness-to pay thresholds.

Figure 12: Cost Effectiveness Acceptability Curves (CEAC) Based on Descriptive Statistics



Note: Y axis reflect probability of the therapy to be cost-effective at different willingness-to-pay values denoted in EUR (X axis).

Figure 13: Cost Effectiveness Acceptability Curves (CEAC) Based on HM 1 Without Covariates

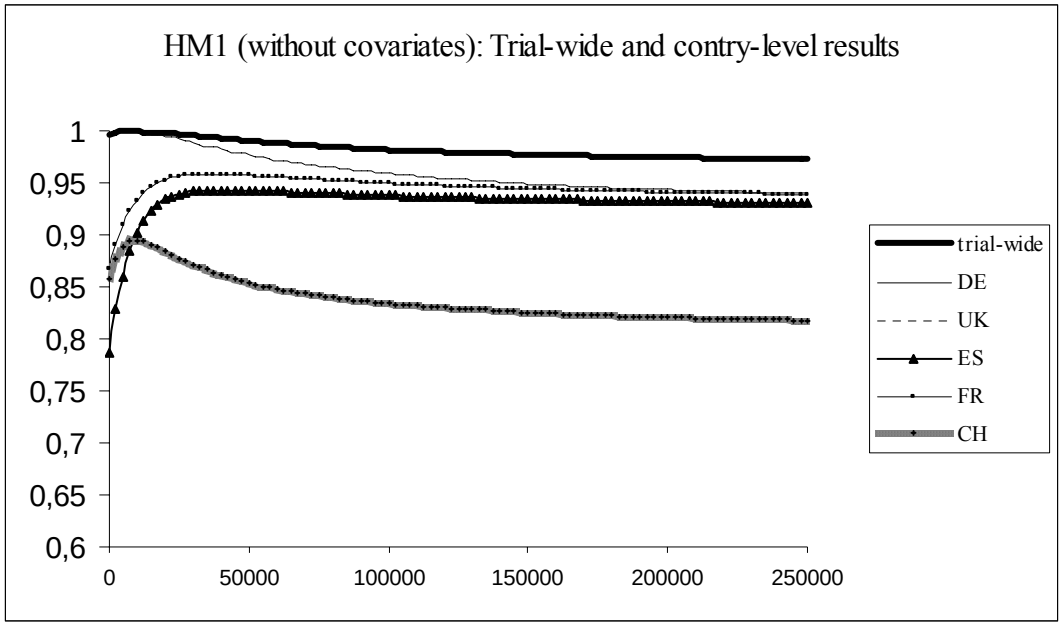


Figure 14: Cost Effectiveness Acceptability Curves (CEAC) Based on HM 2 With Patient- and Country-Level Covariates

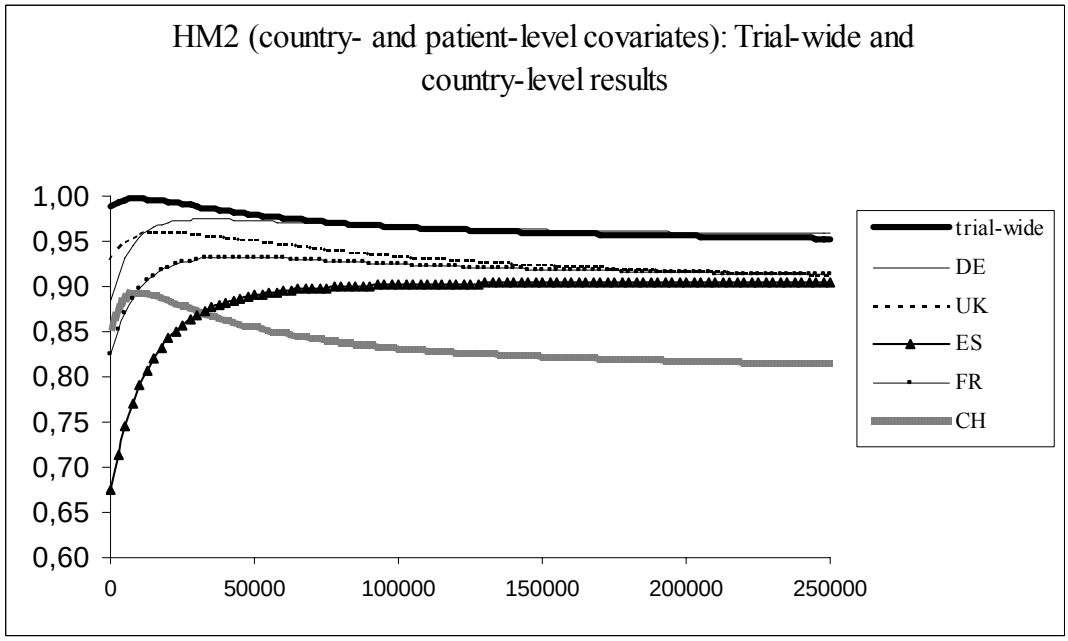
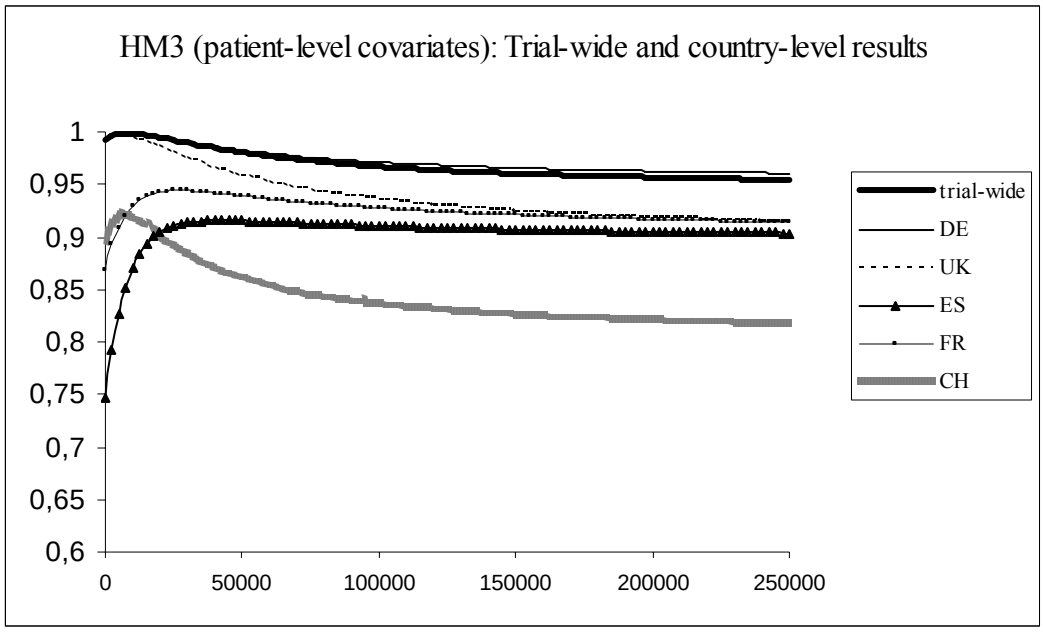


Figure 15: Cost Effectiveness Acceptability Curves (CEAC) Based on HM 3 With Patient-Level Covariates



4.3.5 Country-level LOS Estimates While Accounting for Data Hierarchy and Heterogeneity

Objective 10: Calculation of trial-wide and country-level estimates of the incremental LOS using patient-level data and accounting for within country clustering.

The results of HM 4 in Table 22 suggest considerable variation of incremental LOS between 10 countries. Trial-wide estimate shows reduction in LOS for the patients receiving treatment comparing to standard therapy (-0.56 CR -1.35 – 0.29). This result is not consistent with country-level estimates in all countries: in Switzerland, Israel, Russia and Greece incremental LOS is in favor of comparator. The estimates for Russia (7.96 CR -1.97 – 28.75) are considerably different from other countries and trial-wide results. Also, there is a high uncertainty around the estimate reflected in wide credibility intervals. Small sample size of Russian sub-group (12 patients in both arms) does not allow for robust sub-group analysis, therefore the high difference between treatment and comparator arm just by chance can not be ruled out. Owing to small sample size and high uncertainty of the estimate, the results from Russia do not have major impact on

trial-wide estimates (confirmed in model simulation without Russian patients; results are not presented here).

Table 22: Results of Hierarchical Model 4 for Incremental LOS without Covariates

Setting	Mean	SD	MC error	2.50%	Median	97.50%
Trial-wide	-0.56	0.43	0.01	-1.35	-0.57	0.29
Germany	-1.70	1.80	0.02	-5.21	-1.71	1.93
Spain	-1.39	1.62	0.02	-4.57	-1.41	1.85
UK	-1.32	1.02	0.01	-3.30	-1.32	0.69
France	-0.88	0.73	0.01	-2.32	-0.88	0.56
Switzerland	0.09	1.44	0.02	-2.70	0.06	3.00
Belgium	-2.29	0.99	0.01	-4.27	-2.28	-0.36
Greece	0.51	0.73	0.01	-0.92	0.51	1.96
Israel	0.74	1.58	0.02	-2.28	0.67	4.00
Russia	7.96	10.06	0.23	-1.97	6.00	28.75
South Africa	-0.25	0.51	0.01	-1.25	-0.25	0.74

Objective 11: Calculation of trial-wide and country-level estimates of incremental LOS using patient-level data and accounting for within country clustering while controlling for patient-level and country-level covariates.

Similar to the cost-effectiveness analysis with HM 2, we included full set of country- and patient-level covariates in HM 5. Trial-wide estimate of incremental LOS suggest shorter hospitalization for the patients receiving treatment (-0.65 CR -1.40 – 0.13, see Table 23). Again, similar to results of HM 4, there is considerable variation in country-level estimates: in Spain, UK, France, Belgium, Greece and South Africa patients in the treatment arm spent less time in the hospitals, whereas in Germany, Switzerland, Israel and Russia the patients receiving comparator have no substantial difference or shorter LOS. Inclusion of covariates considerably increased uncertainty of the estimates as reflected in wide credibility intervals in HM 5.

Table 23: Results of Hierarchical Model 5 with Patient- and Country-Level Covariates

Setting	Mean	SD	MC error	2.50%	Median	97.50%
Trial-wide	-0.65	0.39	0.00	-1.40	-0.66	0.13
Germany	-0.10	25.47	0.60	-50.09	-0.38	47.71
Spain	-0.73	21.05	0.50	-40.43	-0.61	40.17
UK	-2.51	19.68	0.46	-40.29	-2.82	35.16
France	-1.20	17.97	0.42	-35.61	-0.94	33.55
Switzerland	0.27	14.33	0.33	-29.38	1.33	26.95
Belgium	-1.26	6.82	0.16	-13.99	-1.33	12.34
Greece	-1.21	13.38	0.32	-25.92	-1.14	25.07
Israel	0.20	4.97	0.11	-8.87	-0.05	10.55
Russia	6.55	47.56	1.10	-86.76	7.56	98.05
South Africa	-1.28	15.51	0.37	-30.94	-1.32	28.44

HM 6 is calculated using patient-level covariates only. Trial-wide estimate of incremental LOS suggest shorter LOS for the patients receiving treatment (-1.24 CR- 2.20 – -0.25, see Table 24). Similar to results of HM 4 and HM 5, there is variation of country-level estimates: in Germany, Spain, UK, France, Belgium, Greece, Switzerland and South Africa patients in treatment arm spent less time in the hospitals, whereas in Israel and Russia the patients on comparator have shorter LOS. The precision of the country-level estimates is improved comparing the credibility intervals in HM5.

Table 24: Results of Hierarchical Model 6 with Patient-Level Covariates

Setting	Mean	SD	MC error	2.50%	Median	97.50%
Trial-wide	-1.24	0.50	0.00	-2.20	-1.25	-0.25
Germany	-1.46	0.95	0.01	-3.32	-1.46	0.39
Spain	-0.55	1.33	0.01	-3.14	-0.55	2.10
UK	-2.79	0.84	0.01	-4.47	-2.79	-1.13
France	-1.62	1.45	0.01	-4.46	-1.63	1.29
Switzerland	-0.51	1.45	0.01	-3.30	-0.54	2.44
Belgium	-1.47	1.51	0.01	-4.44	-1.48	1.57
Greece	-1.03	0.63	0.00	-2.27	-1.02	0.18
Israel	0.10	0.79	0.00	-1.47	0.10	1.66
Russia	7.52	8.69	0.07	-1.86	5.76	27.50
South Africa	-1.57	0.70	0.01	-2.93	-1.58	-0.18

Objectives 12: Overall Comparison and Assessment of model fit after country- and patient- level covariates adjustment.

Table 26 provides the overview of the results as calculated by HMs 4 – 6 and descriptive statistics. Table 25 summarizes results for DIC for models 4-6. Figure 16 and Figure 17 provide visual presentation of incremental LOS obtained with hierarchical models and descriptive statistics.

Data fit of the models is measured by DIC. Model with lower DIC and change in amount of 5-10 points is regarded as improvement of data fit (Spiegelhalter et al., 2002; The BUGs Project, 2010). DIC for HM 4 is 3434.6, HMs 5 and 6 maintained better data fit confirmed by a lower DIC of 3354.34 and 3353.99 respectively. Comparing HM 5 and HM 6, we state, that HM 6 provides better precision of the mean incremental costs estimates as showed by narrower credibility intervals.

Comparison of results obtained with descriptive statistic and hierarchical models indicates that different methods may lead to different country-level conclusions. For instance, in Spain, Switzerland, Belgium, Israel, Russia and South Africa the mean descriptive values suggest that new treatment has either no substantial effect on LOS or even prolongs LOS comparing to standard treatment. Mean incremental LOS, as calculated by HM 6, suggest that in all the above countries but Russia and Israel new treatment reduces LOS. Country-level results calculated with hierarchical models 4-6 are shrunk toward the trial-wide results. Inclusion of patient-level and country-level covariates has an impact on both mean and incremental trial-wide and country-level LOS estimates. Similar to HM2, inclusion of country-level covariates into HM 5 increased uncertainty of the estimates as reflected in wider credibility intervals. Results for Russian setting suggest some unique features which could not be fully explored due to limited sample size. The patient from this setting appeared to be substantially younger (35.58, SD 12.60) than trial average (55.57, SD 20.03); were heavy smokers (12.58, SD 12.56) comparing to the trial-wide (5.81, SD 11.01), but had less patients with severe CAP (0.33, SD 0.49, trial-wide: 0.52, SD 0.50). In this case, trial-wide results as calculated by hierarchical model may be considered for the country-level decision-making, since country-level estimates may be susceptible for bias due to small sample size.

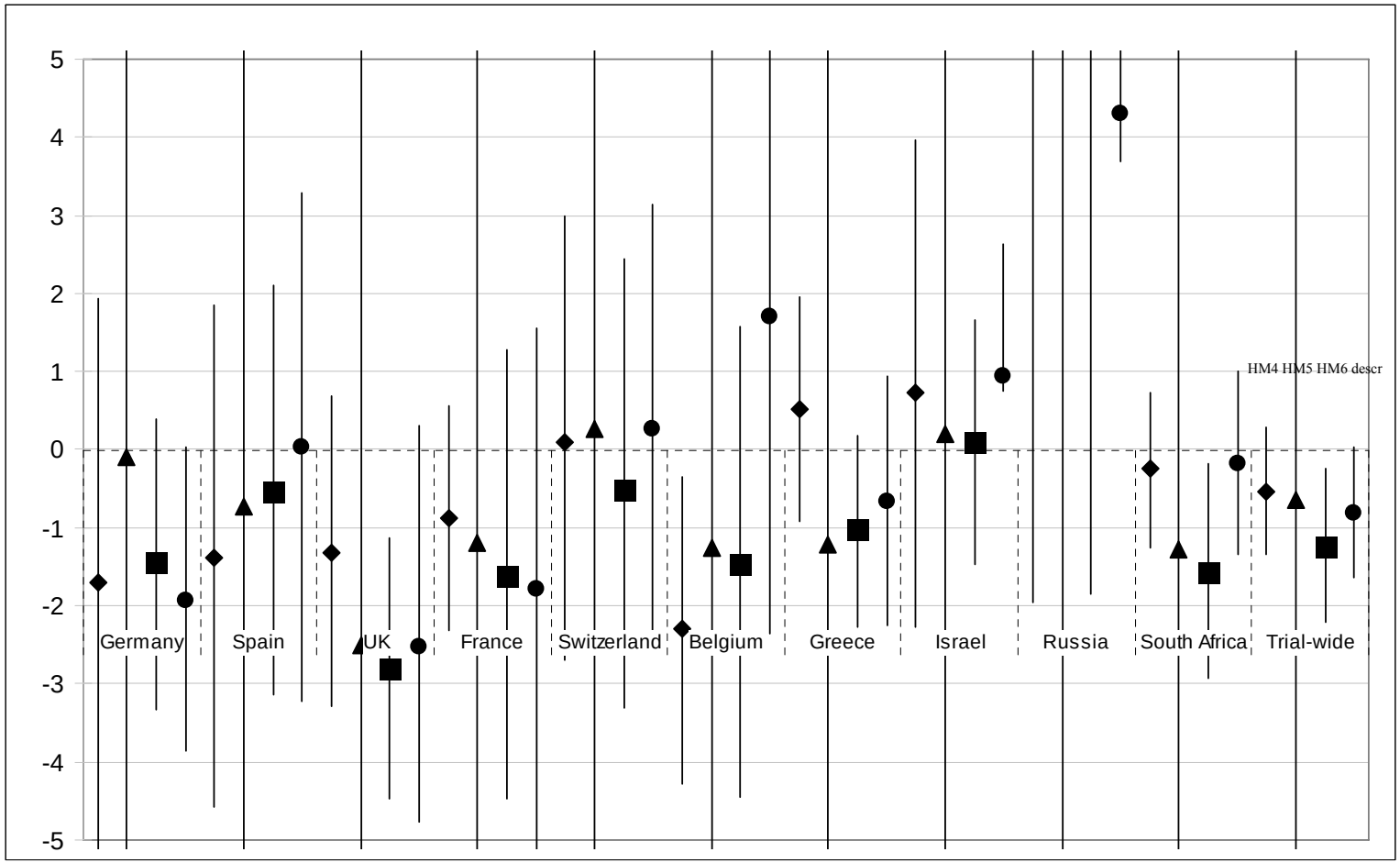
Table 25: Deviance Information Criteria (DIC) for Hierarchical Models 4-6

Model	HM 4 without covariates	HM 5 all covariates	HM 6 pat-level covariates
DIC	3434.62	3354.34	3353.99

Table 26: Results of Incremental LOS Estimated with Hierarchical Models 4-6 and Descriptive Statistics

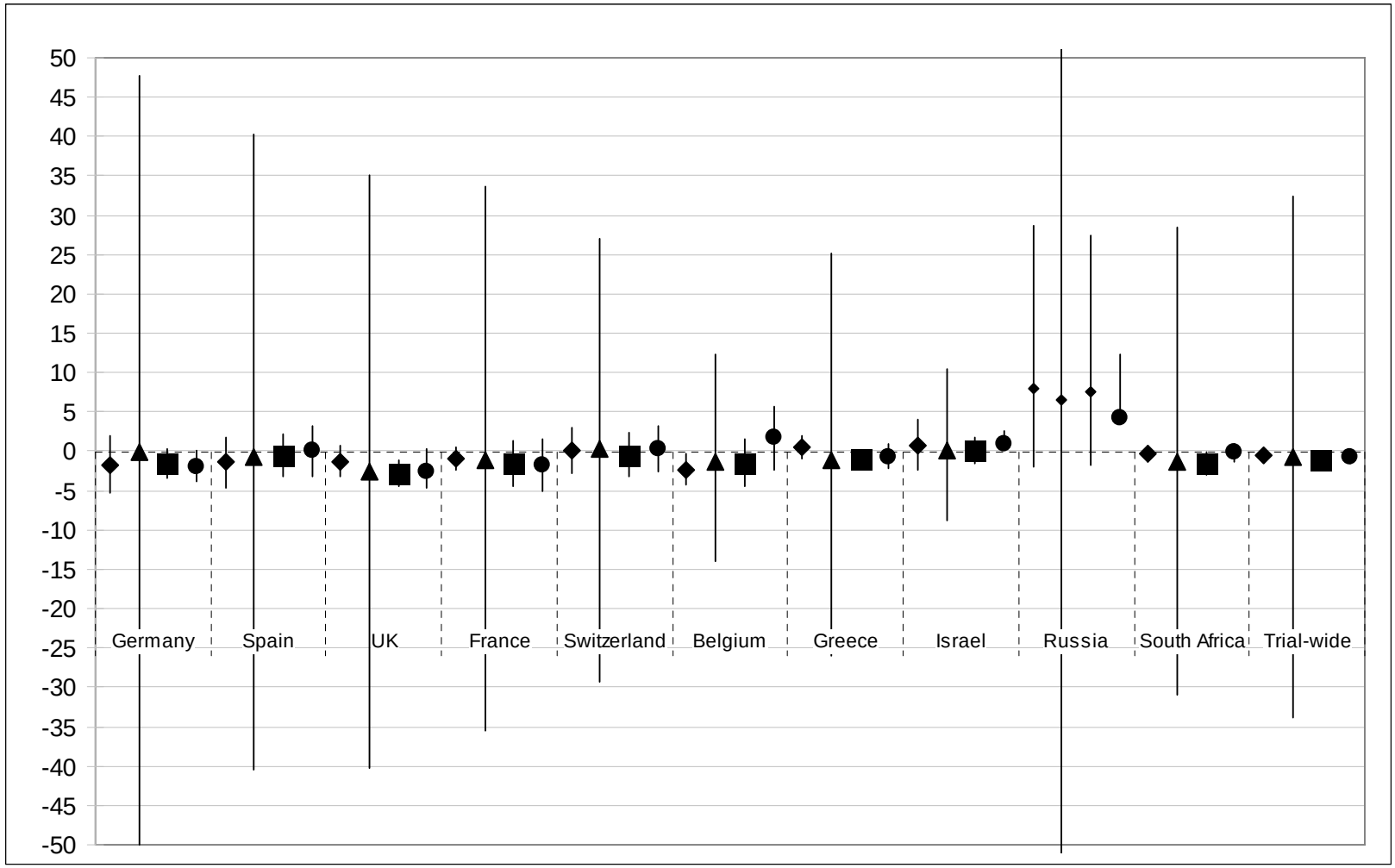
Incremental LOS	HM4: 10 CNTR LOS			HM5: 10 CNTR LOS ALL COV			HM6: 10 CNTR PAT COV			DESCR STAT		
	mean	CR 2.5	CR 97.5	mean	CR 2.5	CR 97.5	mean	CR 2.5	CR 97.5	mean	CR 2.5	CR 97.5
Germany	-1.70	-5.21	1.93	-0.10	-50.09	47.71	-1.46	-3.32	0.39	-1.94	-3.85	0.04
Spain	-1.39	-4.57	1.85	-0.73	-40.43	40.17	-0.55	-3.14	2.10	0.03	-3.22	3.28
UK	-1.32	-3.30	0.69	-2.51	-40.29	35.16	-2.79	-4.47	-1.13	-2.53	-4.76	0.30
France	-0.88	-2.32	0.56	-1.20	-35.61	33.55	-1.62	-4.46	1.29	-1.78	-5.12	1.56
Switzerland	0.09	-2.70	3.00	0.27	-29.38	26.95	-0.51	-3.30	2.44	0.26	-2.61	3.13
Belgium	-2.29	-4.27	-0.36	-1.26	-13.99	12.34	-1.47	-4.44	1.57	1.70	-2.35	5.75
Greece	0.51	-0.92	1.96	-1.21	-25.92	25.07	-1.03	-2.27	0.18	-0.66	-2.26	0.94
Israel	0.74	-2.28	3.97	0.20	-8.87	10.55	0.10	-1.47	1.66	0.95	0.74	2.64
Russia	7.96	-1.97	28.75	6.55	-86.76	98.05	7.52	-1.86	27.50	4.31	3.69	12.32
South Africa	-0.25	-1.25	0.74	-1.28	-30.94	28.44	-1.57	-2.93	-0.18	-0.17	-1.34	1.00
Trial-wide	-0.56	-1.35	0.29	-0.65	-33.75	32.47	-1.24	-2.20	-0.25	-0.81	-1.64	0.03
DIC	3434.62			3354.34			3353.99			NA		

Figure 16: Results of Incremental LOS Estimated with Hierarchical Models 4-6 and Crude Estimates (high definition)



Note: Incremental LOS is expressed in days (Y axis)

Figure 17: Results of Incremental LOS Estimated with Hierarchical Models 4-6 and Crude Estimates (low definition)



4.3.6 Use of Covariates for External Generalizability

Objective 13: Assessment of covariates effects on mean and incremental costs and effectiveness estimates and possibility to generalize results of the study to non-study settings on hand of explored covariates (external generalizability).

Table 21 summarizes results of hierarchical models 1 – 3 and crude estimates; the covariates coefficients from models 2 and 3 are presented in Table 27 and Table 28 respectively. Effect of a covariate is splitted into direct effect on mean and interaction effect on incremental costs and LOS. Covariates coefficients have partial effect on the outcome of interest, meaning that increase of covariate in one unit will increase outcome of interest for an amount of coefficient holding all other parameters constant. Positive interaction effect increases mean difference in costs or effectiveness by a regression coefficient for a unit of covariate above the trial-wide average. For example, mean incremental cost are increased by 4.67 per year of age above the trial-wide mean age. In the case mean incremental effect is negative – therapy costs of treatment are lower than the costs of comparator – incremental costs will be lower in absolute value.

As mentioned above, inclusion of covariates improved data fit measured by reduced DIC when comparing models 1, 2 and 3. We have gain in precision of estimates reflected in narrower credibility intervals for estimates in model 3 comparing to model 2.

Trial-wide incremental cost are negative as estimated by hierarchical models and descriptive statistics: -555.50 (95 CR -945.00 – -160.10) for HM 1, -475.10 (95 CR -869.00 – -81.84) for HM 2, -472.30 (95 CR -839.00 – -100.00) for HM 3 and -659.69 (CI -1305.51 – -13.78) based on crude estimate. Incremental costs in models 2 and 3 are clearly affected by inclusion of covariates. Overall, country-level covariates (GDP-total health expenditures as proportion of GDP, HOSP – number of hospital beds per 100 000 inhabitants, RES – antibiotics resistance) had small impact on incremental costs. This evidence suggests that it might be difficult to identify "right" contextual (country-level) covariate with sufficiently strong causal association with parameters of interest in order to predict incremental effect of the therapy based on country-level parameters alone. However, patient-level covariates (AGE – patient's age at admission, SMOK – mean number of cigarettes smoked per day, CAP – severity of CAP condition at the hospital admission) showed different picture. Covariates age, number of cigarettes

smoked per day and severity of the condition at the hospital admission have effect on mean costs. Patients' age and severity have positive effect on incremental costs, whereas effect of age is more pronounced than the effect of severity of the condition. Smoking had small negative effect on incremental costs.

Table 27: Covariates Coefficients from Hierarchical Model 2

	Coefficient		Mean	CR 2.5	CR 97.5
Interaction effect	CAP	omegaCX2	1.47	-18.00	21.01
	RES	omegaCX3	-0.07	-19.57	19.50
	AGE	omegaCX4	4.51	-7.44	16.23
	GDP	omegaCY1	0.05	-19.55	19.72
	HOSP	omegaCY2	-0.26	-4.52	3.50
	SMOK	omegaCY3	-0.34	-15.43	15.28
	AGE	omegaEX4	0.00	0.00	0.00
Direct effect	CAP	thetaCX2	2.43	-17.10	21.85
	RES	thetaCX3	0.10	-19.47	19.76
	AGE	thetaCX4	18.16	8.09	28.32
	GDP	thetaCY1	0.07	-19.64	19.67
	HOSP	thetaCY2	1.76	-7.25	9.37
	SMOK	thetaCY3	12.76	-0.20	25.85
	AGE	thetaEX4	0.00	-0.01	0.00

Note: Patient-level covariates CAP denotes severity of condition at baseline, AGE – patients' age; SMOK – mean number of cig/day; country-level covariates RES denotes antibiotics resistance, GDP – total health expenditure in % of GDP, HOSP – number of hospital beds per 100 000.

Table 28: Covariates Coefficients from Hierarchical Model 3

	Coefficient		Mean	CR 2.5	CR 97.5
Interaction effect	CAP	omegaCX2	1.45	-17.83	20.82
	AGE	omegaCX4	4.67	-7.03	16.46
	SMOK	omegaCY3	-0.22	-15.49	15.44
	AGE	omegaEX4	0.00	0.00	0.00
Direct effect	CAP	thetaCX2	2.45	-16.79	21.90
	AGE	thetaCX4	18.23	8.15	28.24
	SMOK	thetaCY3	12.80	-0.13	25.95
	AGE	thetaEX4	0.00	-0.01	0.00

Objectives 14: Assessment of covariates effects on mean and incremental LOS estimates and possibility to generalize results of the study to non-study settings on hand of explored covariates (external generalizability).

Covariates coefficients calculated in models 5 and 6 are summarized in Table 29 and Table 30.

Inclusion of covariate improved data fit when comparing DIC of model 4 (3434.62) with DIC of models 5 and 6 (3354.34 and 3353.99 respectively). Estimated trial-wide incremental LOS is comparable between hierarchical models and crude estimate: -0.56 (95 CR -1.35 – 0.29) for HM 4, -0.65 (95 Cr -33.75 – 32.47) for HM 5, -1.24 (95 CR -2.20 – -0.25) for HM 6 and -0.81 (CI -1.64 – 0.03) based on crude estimate. Incremental LOS in models 5 and 6 are clearly affected by inclusion of covariates.

Inclusion of country-level covariates (GDP- total health expenditures as proportion of GDP, HOSP – number of hospital beds per 100 000 inhabitants, RES – antibiotic resistance) introduced high uncertainty into parameters' estimation, what is reflected in both credibility intervals of covariates coefficients and incremental LOS results in HM 5. Covariate GDP and resistance have small direct effect on mean LOS. In contrast, number of hospital beds per 100 000 inhabitants have pronounced direct effect,

suggesting that in countries with number of hospital beds higher than trial-average, the length of hospitalization is longer. The interaction effects suggest very small impact of country-level covariates on the LOS as well as high uncertainty of estimates, as reflected by wide credibility intervals. In general, wide credibility intervals are reflecting lack of information to quantify association of country-level parameters and incremental LOS partially attributed to a limited sample size.

After removing of country-level covariates, we have gained precision of estimates reflected in narrower credibility intervals for Germany, Spain, Switzerland, Israel and Russia for estimates in HM 6 comparing to models 4 and 5.

Similar to HM 2 and 3, direct and interaction effects of patient-level covariates are more precisely estimated comparing to country-level coefficients (narrower credibility intervals).

Table 29: Covariates Coefficients from Hierarchical Model 5

	Coefficient		Mean	CR 2.5	CR 97.5
Interaction effect	CAP	omegaCX2	0.90	-0.19	2.00
	RES	omegaCX3	1.01	-10.63	13.00
	AGE	omegaCX4	0.00	-0.03	0.03
	GDP	omegaCY1	-0.74	-19.68	18.09
	HOSP	omegaCY2	-0.86	-18.29	16.28
	SMOK	omegaCY3	-0.03	-0.08	0.03
Direct effect	CAP	thetaCX2	0.93	0.17	1.69
	RES	thetaCX3	-0.27	-3.60	3.09
	AGE	thetaCX4	0.06	0.04	0.09
	GDP	thetaCY1	0.86	-8.13	8.95
	HOSP	thetaCY2	6.71	-0.73	13.89
	SMOK	thetaCY3	0.04	0.00	0.08

Table 30: Covariates Coefficients from Hierarchical Model 6

	Coefficient		Mean	CR 2.5	CR 97.5
Interaction effect	CAP	omegaCX2	0.89	-0.21	1.98
	AGE	omegaCX4	0.00	-0.03	0.03
	SMOK	omegaCY3	-0.03	-0.08	0.03
Direct effect	CAP	thetaCX2	0.95	0.18	1.70
	AGE	thetaCX4	0.06	0.04	0.09
	SMOK	thetaCY3	0.04	0.00	0.08

The effect of covariates on mean and incremental estimates is similar in HM 2 and 3 and HM 5 and 6: country-level covariates have effect on mean costs and LOS, but small und uncertain impact on incremental estimates. In the countries with the parameters above the trial average - GDP expenditures, hospitals beds per capita, or with higher antibiotic resistance - mean average costs will tend to be higher. Analysis of full dataset with 10 countries showed that in particular number of hospital beds per capita is associated with increased mean LOS in the countries with hospital capacity above trial-average. Patient-level covariates have effect on both mean and incremental estimates. In absolute terms the incremental difference between elderly patients and patient with severe CAP is diminishing in both incremental costs and LOS analysis; for the smokers the incremental effect will increase with amount of cigarettes smoked.

These findings resonate with clinical evidence about CAP indication and treatment drug. Treatment has targeted activity against Streptococcus pneumoniae (including multi-drug resistant strains), Haemophilus influenzae, Moraxella catarrhalis, Mycoplasma pneumoniae, and Chlamydia pneumoniae. In vitro studies have demonstrated that treatment achieves drug concentrations at least three times higher in lung tissues than the older compounds: fluoroquinolones, levofloxacin or ciprofloxacin (Mandell et al., 2004). Clinical studies have demonstrated that treatment monotherapy is as effective (Katz et al., 2004; File et al., 2001; File, 2003) or more effective (Finch et

al., 2002; Drummond et al., 2003) than comparator combination regimens, usually containing a β-lactam and/or macrolide. Treatment (IV, 400 mg) followed by p.o. (400 mg) for 7-14 days had superior clinical success rates (93% vs 85%) and superior bacterial eradication (94% vs 82%), and treatment-treated patients were afebrile in a shorter time (by day 2, 59% vs 47%) than comparator-treated patients (IV co-amoxiclav, 1.2 g t.i.d., followed by p.o. co-amoxiclav, 625 mg, t.i.d., with or without IV/p.o. clarithromycin, 500 mg b.i.d. for 7-14 days) in 628 adult patients with CAP requiring initial parenteral therapy (Finch et al., 2002). A faster IV to oral switch therapy by day 5 was observed for patients receiving treatment than comparator in severe CAP (n = 376) (Lode et al., 2003). Thus, clinically proven safety and tolerability profile of the treatment drug supports our findings on cost and LOS savings comparing to the standard therapy in general population. However, the incremental difference is smaller for elder and severe CAP patients. The evidence suggests that elder patients with CAP are at higher mortality and re-admission risk. Severely sick and /or elder patients recover more slowly due to existing co-morbidities (Garcia-Ondonez et al., 2001). Thus, treatment outcomes and decision about the discharge (and thus savings of LOS and overall costs) may not rely on the efficacy of the therapy alone, but rather by the patient's overall functional status, need for care and rehabilitation and destination of discharge (e.g. nursing home or family care).

Interesting finding about potentially increased incremental difference in LOS and costs for smokers finds supporting evidence in the published literature. Smoking effects multiple immune mechanisms in lungs in a complex way. The exact mechanism of action is not yet well understood (Robbins et al., 2006). Hand and colleagues (1985) investigated antibiotics uptake in smokers. Their findings suggest that lipid soluble antibiotics achieved high concentration in the pulmonary alveolar macrophages cells in contrast to poorly uptake of lipid insoluble antibiotics (Hand et al., 1985). In fact, treatment is a fluoroquinolone and is lipid-soluble; the comparator is a B-lactam and has low solubility in lipids. Certainly, relationship between smoking and immune response is very complex and differ across the age group and co-morbidity status. Thus, the effect of new therapy in smokers and potential savings merit further investigation in larger settings.

4.4 Summary of Chapter 4

This section comprises the results of the case study. Mean and incremental costs and effectiveness were estimated for UK, France, Germany, Spain and Switzerland using crude country-level sub-samples and hierarchical models with and without covariates. Mean and incremental duration of hospitalization was estimated for the full 10-country set (UK, France, Germany, Spain, Switzerland, Israel, Russia, Belgium, South Africa and Greece) using crude country-level sub-samples and hierarchical models with and without covariates. Between-country heterogeneity was explored by using qualitative and quantitative homogeneity test.

Descriptive statistics using country-level sub-groups indicated potential between-country heterogeneity.

New treatment is cost saving in Germany, UK, France and Switzerland, but not in Spain. The magnitude of cost saving is strongly pronounced in France and UK, less so in Switzerland, whereas the data for Spain suggests the comparator is cost saving over the treatment. With regard to incremental effectiveness, treatment is more efficacious in Germany, Spain, UK, France, Greece, Israel and Russia, but has no pronounced difference or comparator is more efficacious in South Africa, Belgium and Switzerland. Similarly, duration of hospitalization is shorter under treatment in Germany, UK, France, Greece and South Africa and longer in Spain, Switzerland, Belgium and Israel. Overall, country-level results indicate potential between-country heterogeneity.

Homogeneity test is not conclusive for decision about between-country heterogeneity.

Results of qualitative homogeneity tests suggest absence of country-by-treatment interactions in incremental effectiveness, length of stay, costs and INMB. That indicates essentially, that the effect's direction appears to be homogeneous in all countries, i.e. new treatment is more efficacious, saves hospitalization time and costs. However, test power for incremental effectiveness, cost, INMB and LOS (5 countries) parameters is low. Particularly low test power is seen for incremental costs and INMB, where the variability of estimates is high, and thus, the heterogeneity may remain undetected.

Test for quantitative interactions, similar to the results of the test for qualitative interaction, suggests absence of between-country heterogeneity for incremental

effectiveness, costs and INMB. However, the results for incremental LOS indicate significant quantitative between-country heterogeneity, suggesting significant difference in the magnitude to which the treatment reduces length of hospitalization in different countries.

Overall, it remains unclear, to which extend could trial-wide cost, effectiveness and resource use estimates be considered informative for the decision making in individual countries.

Hierarchical models are used as compromise between country-level and trial-wide analysis for the calculation of trial-wide and country-level estimates. Inclusion of covariates improves model fit.

Mean and incremental costs and effectiveness were calculated using bivariate hierarchical models for effectiveness distributed with Bernoulli and costs with gamma distributions to account for nesting of the data within individual countries and right-skewness of the costs distribution.

Patient-level covariates (age at admission, severity of CAP condition at admission and mean number of cigarettes smoked per day) and country level covariates (total health expenditures as proportion of GDP, number of hospital beds per 100 000 inhabitants, antibiotic resistance) covariates improved data fit in hierarchical models with costs and LOS comparing to the hierarchical models without covariates as confirmed by lower DIC.

Country-level covariates have limited predictive ability for incremental effects to generalize case-study results beyond the setting of participating countries.

Country-level covariates had very small impact on incremental costs estimates. Patient-level covariates (age, number of cigarettes smoked per day and severity of the condition at the hospital admission) have effect on mean costs. Patients' age and severity have positive effect on incremental costs, whereas effect of age is more pronounced than the effect of severity of the condition.

Similar to costs estimation, covariate GDP and resistance have very small direct effect on mean LOS. In contrast, number of hospital beds per 100 000 inhabitants have pronounced direct effect, suggesting that in countries with number of hospital beds higher than trial-average, the length of hospitalization is longer. The interaction effects

suggests very small impact of country-level covariates on the LOS as well as high uncertainty of estimates, as reflected by wide credibility intervals. Severity of condition at admission, measured at patient-level, had direct and interaction effect on LOS. Thus, patients with severe condition were more likely to stay in the hospital longer. Positive interaction effect suggests that incremental difference between treatment and comparator will be smaller for the patients with severe condition. In general, it may be difficult to predict incremental costs and effectiveness based on explicit consideration of macro country-level parameters only.

5 Discussion

Summary of thesis objectives and major findings

Between-country variability in clinical and economic parameters in health economic studies is broadly acknowledged in the literature. However, assessment and quantification of heterogeneity and it implication for generalizability of cost-effectiveness estimates has not been extensively investigated.

In this work we explored approaches to improve generalizability of multinational patient-level studies by addressing two main objectives:

1) Review of published methods used in multinational health economic studies to explore heterogeneity and to improve generalizability.

2) Application of selected methods to the multinational patient-level dataset in a case study to evaluate between-country variability in cost, resource use and effectiveness parameters, to identify potential covariates explaining between-country heterogeneity and to calculate country-level and trial-wide cost, effectiveness and resource use estimates.

First objective was addressed by conducting a targeted review of health economic literature and health economic guidelines. Essential findings of the literature review are the following:

- Between-country variability in cost-effectiveness and resource use estimates could be attributed to wide geographical differences in patient-level and country-level (or setting-level) parameters.
- Studying of between-country variability is not yet established part of health economic analysis. It is recognized, however, that the estimates based on the aggregated analysis (e.g. trial-wide resource use) may not be appropriate for the country-level decision making in all participating countries.

- Homogeneity test and hierarchical modelling emerged as analytical strategies to address heterogeneity and appropriate handling of hierarchical data for calculation of trial-wide and country-level parameters.

Second objective was attained by a case study, which illustrated application of homogeneity test and hierarchical models. The major results are the following:

- Qualitative homogeneity test suggested no evidence for between country heterogeneity in incremental clinical and economic effects. Quantitative homogeneity test revealed between-country heterogeneity in incremental length of hospitalization.

- Limited test power for qualitative interactions and potential quantitative heterogeneity does not suggest concordant evidence for between-country homogeneity and hence feasibility of pooling data across participated countries.

- Trial-wide and country-level estimates of incremental effectiveness, costs and LOS were calculated by using bivariate hierarchical models with and without covariates. Hierarchical modelling accounts for clustered data within individual countries as well as allow for using appropriate not-normal distribution for the parameters of interest.

- Using of covariates substantially improved model fit. Patient-level covariates had effect on mean and incremental cost and LOS. Country-level covariates had limited effect on incremental economic estimates. The generalizability of the study findings beyond the study setting may not be feasible by explicit considerations of country-level covariates alone.

Relevance of the thesis research subject to Public Health realm

Health economic studies, being part of HTA, are vehicles to collect clinical and economic evidence to inform health political decision-making. Facing the technological revolution and emerging of new expensive technologies, efficient allocation of limited resources is essential to maintain access to health technologies and to provide health care services to the broad population. The “invisible hand” of the competitive market will not necessarily allocate limited resources to get maximum health gain. Free market failure is caused by asymmetric information, uncertainties and externalities. Inefficiency of health care market should be corrected by governmental institutions and policy. In

order to improve health care system performance (and not to do more harm than good, since competitive market incentives are applicable for many health care sectors), political decision-making should be based on good evidence. In the framework of HTA the evidence about the impact of new technology is collected and evaluated from the broad societal perspective. The aim of HTA is to inform decision makers and to promote rational use of health care services. Economic analysis of new technology is one of the aspects of HTA. Evidence on economic impact of technologies (on costs and outcomes) is limited and is often available from international studies only. Development of generalizability methods will be beneficial for both health policy and research for two main reasons:

1) Generalizability methods will improve data utilization from available sources and conducted economic evaluations and provide local tailored evidence for the decision-making in particular setting.

2) Assessment of between-country heterogeneity will contribute to cross-country comparison, understanding the factors inducing the differences between the settings and indicate the more efficient way of health provision and production.

Theoretical considerations for studying between-country heterogeneity and generalizability

Discussion about between-country differences in mean and incremental resource use and cost is mainly driven by empirical considerations in differences of health-care inputs (e.g. proportion of GDP spent on health care, number of hospital beds per capita etc.) and outputs (mortality, morbidity, resource use). The empirical approach alone may fail to uncover the whole complexity of the relationship between different clinical and wider socio-economic factors operating at various levels (patient-, formulary-, regional-, country-levels) and influencing mean and incremental clinical and economic outcomes. Thus, normative theoretical considerations are important to provide guidance for empirical analysis.

In this thesis we outlined major economic theories related to health policy and empirical research and linked these to Public Health realm. Welfare economic theory advocates political intervention and regulation of health care market. In fact, health care production is strongly influenced by health policy, institutions and setting in which they are operating. From the theoretical perspective New Institutional Economics provides a

conceptual and theoretical basis for studying policy, interrelationship between institutions, incentives and efficiency in health care markets. New Institutional Economics, being at it infancy stage, is yet to be broadly applied in studying of efficiency and incentives of different institutions and political interventions and thus, expand our understanding of contextual effects on health care. Exploring of the contextual effect will enable researches to account for between-country heterogeneity and factor this into quantitative models to improve generalizability of multinational health economic studies.

Exploring heterogeneity in patient-level studies using homogeneity test

Homogeneity test has been suggested to detect between-country heterogeneity in effect's direction (qualitative interaction) and magnitude (quantitative interaction). In the absence of qualitative interaction in the effect direction, pooled estimates are considered applicable for the decision-making in individual countries participated in the trial (Cook et al., 2003).

Homogeneity test can be easily implemented and could be routinely used as part of health economic analysis. However, the conclusions about the absence of heterogeneity may not be drawn based on the results of the test alone due to the following limitations:

- Low test power may not allow to detect qualitative interaction in economic parameters, when it actually exists.
- Impossibility to formally explore and account for factors inducing heterogeneity. It is possible, however, to group countries with no evidence for treatment-by-country interaction.
- Unclear extend to which the trial-wide results, in absence of qualitative interactions, should be considered applicable for the decision-making in individual countries. Using pooled estimates for all participating countries would imply the high degree of similarity between geographical settings, what could be unrealistic and difficult to prove.

In addition, it has been argued, that statistical significance is not primarily relevant in health economic analysis (Claxton, 1999). In fact, in many cases it might be plausible to assume between-country heterogeneity and handle analysis with appropriate methods, such as hierarchical modelling.

Calculation of trial-wide and country-level estimates using hierarchical models (internal generalizability)

Evolving health economic literature demonstrated application of hierarchical modelling for calculation of setting-specific estimates using patient-level data collected in multinational trials (Rice & Jones, 1997; Manca et al., 2005; Grieve et al. 2005; 2007; Pinto et al., 2005; Willan & Kowgier, 2007; Manca et al., 2007).

Hierarchical models provide a flexible framework to address the generalizability of the estimates across and possibly beyond the study jurisdictions. It allows calculation of country-level resource use, costs and cost-effectiveness by a mean of the shrinkage estimation - a weighted average of country-level and trial-wide (pooled) data. Hierarchical analysis takes into account between- and within settings variability and thus allows for correct reflection of uncertainty around parameter estimates. Moreover, it is possible to apply appropriate not-normal distribution for clinical and economic data. Finally, hierarchical modelling can be used to explore the extent, to which results are influenced by compositional and contextual effects and could be applicable beyond the study setting.

In the case study, we calculated trial-wide and country-level incremental costs and effectiveness estimates using bivariate hierarchical model framework proposed by Nixon and Thompson (2005) and Willan and Kowgier (2007). There are two important advantages of bivariate approach: (1) possibility to model costs and effectiveness separately with appropriate not-normal distributions and (2) include different set of covariates as necessary into costs and effectiveness equations. In our case study we used gamma distribution for costs and Bernoulli for effectiveness estimation and evaluated impact of patient-level and country-level covariates on cost, but not on clinical effectiveness. Clinical effect of the studied drugs was assumed not to be severely influenced by covariates, since pivotal clinical studies are highly controlled "experiments" designed to evaluate efficacy with the minimal risks to patients' health. This means that maximum efforts are applied to restore and preserve health, but probably at different costs for different therapies. This assumption may not be valid for routine clinical practice, where both effectiveness and costs may be influenced by covariates.

Important prerequisites for hierarchical analysis relates to "exchangeability" of data and sufficient sample size.

"Exchangeability" is a crucial assumption for hierarchical analysis and implies that selection of sites and countries participating in the study was carried out randomly and those are representative for the general population, for which the inference is being made. This should be valid for each country and site included into the study. This assumption may not be appropriate for any dataset and specific study design. Different populations may be considered as distinctive, rather than exchangeable, and should be treated with appropriate methodology e.g. using fixed effect model. Here, the shrinkage estimation procedure may not be appropriate. Exchangeability assumption should be considered at the study design stage in specification of patients' and centers' inclusion criteria. As noted by Duncan and colleagues (1998), this assumption is critical, and, similarly to all statistical techniques, hierarchical models are only as good as the data they fit (Duncan et al., 1998).

The second prerequisite relates to the sample sizes. Precise estimation of between- and within-level variation requires sufficient number if units at contextual and compositional levels. Efficient data collection may be achieved by multistage sampling design. Considerations about the optimal sample size are difficult outside the context of specific research question. Hierarchical models are broadly applied in analysis of school performance, where sampling considerations have been discussed. For example, Bryk and Raudenbush recommended sample of 60 students from 160 schools for reliable estimation of school-level random effects, whereas other researchers indicated that smaller group sample and size could be sufficient for specific research questions (Duncan et al. 1998; Bryk & Raudenbush, 1992; Paterson & Goldstein, 1992). Nonetheless, this data requirements may not be feasible for rare clinical conditions or, due to limited funding, many health economic studies will not be able to fulfill these prerequisites, particularly with regards to the number of countries.

The estimation of the hierarchical models may face technical issues in the cases where the number of higher-level settings is insufficient. As Draper (1995) emphasized, broadly used maximum likelihood estimation is implying large sample size. For the cases with small sample of high-level settings, maximum likelihood procedure may provide biased variance estimates and too narrow confidence intervals. The possible way to overcome this problem is to use a fully Bayesian approach (Goldstein, 1995;

Draper, 1995; Gilks et al., 1996; Duncan et al., 1998). This method is implemented in WinBUGS and was used for the model estimation in our case study.

Adjusting for patient- or country-level differences: Hierarchical modelling with covariates (internal generalizability)

Effects of patient- and country-level covariates on economic parameters in hierarchical models are not yet broadly explored in health economic literature (Drummond et al., 2009). Inclusion of covariates into costs and effectiveness equations in hierarchical models can optimize analysis as following:

- Increase precision of the costs and effectiveness estimates.
- Allow for possible case-mix differences in baseline characteristics between treatment and control groups across different countries.
- Adjust for differences in covariates to conform with "exchangeability" assumption in hierarchical models.
- Identify the extent, to which covariate influence incremental costs and effectiveness estimates.

Our case study confirmed that inclusion of covariates into the models improved model fit and reduced uncertainty of the estimates as confirmed by narrower credibility intervals of the respective estimates.

Randomization of patients to different treatment arms should minimize imbalance between treatment and comparator groups in risk and other important parameters; nevertheless, it is appropriate to adjust for covariates to allow for any differences by chance between the groups (Nixon & Thompson, 2005). In addition, controlling for difference in the important clinical parameters between the countries is not secured by randomization procedure. In our case study we adjusted for differences in patients' age, smoking and severity of condition, which showed impact on mean and incremental costs estimates. It means, that variability in resource consumptions and incremental effect of the therapy is partially attributed to differences in patients age, smoking and severity of condition.

Important assumption of hierarchical models is the "exchangeability" of the parameters of interest, since these are drawn from a common distribution (Spiegelhalter

et al., 2002). It implies, that there are no a priori reasons to believe, that estimates are relatively high or relatively low in a particular jurisdiction. This assumption might be critical for multinational trials conducted in countries at different stage of economic development and with different health care coverage. Cost analysis (hierarchical models 1-3) was based on the data from patients coming from Western European countries and treated in all countries using common study protocol. Analysis of LOS (hierarchical models 4-6) used data collected in countries with different level of social security and economic developments, where adjustment of covariates was particularly important.

Inclusion of covariates into analysis allows both to identify "predictors" and to conduct the test for interactions between the groups. In our case study all patient-level covariates showed effect on mean costs and LOS; the covariates "age" (patients age at admission) and "CAP" (severity of condition at admission) showed positive effect on incremental costs and covariate CAP showed effect on incremental LOS.

In the analysis, country-level covariates showed only small effect on incremental cost and LOS estimates. These findings may indicate true absence of the relationship between country-level parameters and incremental costs and LOS. However, it can not be completely ruled out, that our analysis failed to reveal the association between contextual factors due to the following reasons:

- selection bias and controlled phase 3 environment
- ambiguous quantification of covariates
- clustering on the site-level
- assumed linear relationship between the covariates and dependent variables
- measurement scale and extend of dissimilarity.

Inclusion of only few patients in individual countries could lead to a selection bias in country-level samples. For example, the patients from Switzerland were significantly elder than trial-average. As the result, country-level patient sample in the study may not be entirely representative for the country general patient population, and so country-level covariates will fail to explain variability. In our case we attempted to include smoking status as country-level covariate based on WHO data as mean number of cigarettes smoked per person per year (Table 13). As a country-level covariate, smoking

had no effect on mean and incremental costs. We then used smoking as a patient-level covariate measured by mean number of smoked cigarettes per day. By ranking the countries based on cigarette consume, there were no overlap based on these two calculation options (WHO data for mean consume of cigarettes per person per year and study data of mean number of cigarettes consumed per day), implying either potential selection bias of the patients into the study or substantial difference in tobacco consume in patients with CAP and general population.

Quantification of the covariates may be problematic. In the case study we included antibiotic resistance as a country-level covariate. As mentioned earlier, the measurement and quantification of the antibiotic resistance is not unanimous and universally agreed due to clinical complexity of the issue and difficulty to measure this parameter in several countries by the same methodology (EARSS, 2006). Thus, it is possible that incorrect or imprecise assessment of covariates is the reason for it failure to explain variability.

In the case study we assumed two-level hierarchy: the patients were clustered within the countries and selected covariates from respective patient- and country-levels. However, it is well possible, that three-level hierarchical model would be more appropriated, namely: patients nested within trial sites and trial sites nested within the countries. The within-country variation in costs and effectiveness outcomes is well acknowledged in the health economic literature (Cook et al., 2004; Birch & Gafni, 2003). Site-level parameters, e.g. community or teaching university hospital, hospital size, financing structure and others, may have substantial effect on mean and incremental costs and effectiveness. In our case study it was not feasible to conduct three-level hierarchical model with site-level due to small sample size in many centers.

In the case study we modeled the effect of covariates as linear and additive, e.g. a covariate was adding a certain amount to mean and incremental cost or resource use. Covariates may have non-linear effect on the outcomes variable by e.g. multiplying costs by a certain factor. This option should be investigated in the future simulations.

Finally, covariates can be measured on different scale types, e.g. from nominal to ratio scales. The later contains more information and is more sensitive to small differences. This factor is important, when there is relevant, but not pronounced difference in covariates between different countries or settings. In this case, the impact of covariate

on mean and incremental parameters may go undetected due to small sample size. Careful consideration of casual associations between the parameters is always commendable.

Generalizability beyond the study setting: predictive ability of covariates for effectiveness and resource use estimates (external generalizability)

Generalization of the study results to non-study settings, e.g. countries not participating in the specific trials, has relevance for health policy and research. Evaluation of the effects of country-level covariates (contextual effects) on parameters of interest is one possible option to facilitate generalizability beyond the study setting.

The access to longitudinal health, resource use and economic data has forced empirical research to comparisons of various aspects of international health systems, health spending, and outcomes (Retzlav-Roberts et al., 2004). The results of these studies suggest very complex relationship between health spending, outcomes and contextual effects. For example, Babazono and Hillman (1994) stated no association between health outcomes and health care expenditures per capita across 21 OECD countries. Authors suggested that the resource distribution across health and non-health expenditures, and not the total level of funding, have impact on health outcomes. Anderson and colleagues (2003) hypothesized that the difference in health expenditures across OECD countries is forced by high prices of health care services in USA. In addition, they pointed out, that allocation of real resources is critical for health care system efficiency and costs. Or (2000) explored the determinants of population's health status using health production function approach. He identified positive association between health status and expenditures in women and relevance of environmental factors in explaining of prematurely mortality in 21 OECD countries.

In our case study we established association between selected country-level covariates and mean resource use and costs: high antibiotics resistance, number of hospital beds per capita have positive impact on mean costs and LOS. However, country-level covariates showed small effect on incremental costs and resource use, what is the central question for the decision-making about the reimbursement of new therapy. The results of the case study support conclusion of Manca and colleagues (2007), that country-level covariates alone can not be used to predict estimated differences in costs and effects for the countries beyond the study setting. Indeed, finding "good" predictors

for costs and effectiveness parameters is difficult (Nixon & Thompson, 2005; Briggs & Gray, 1999). This issue and selection of relevant covariates out of many possible parameters is particularly important for assessment of impact on incremental differences between the therapies.

Based on results of our case study, we outlined several aspects to consider while selecting the covariates as presented in Table 31. Our case study was based on assessment of the effects of different antibiotic treatments in acute clinical condition. It is plausible to assume, that where new treatment course is not substantially changing current clinical practice comparing to standard treatment, the country-level parameters may not have strong effect on incremental costs and effectiveness. When country-by-treatment interaction is not expected to be strong, use of patient-level covariates may be more appropriate and informative. For example, new treatment may be more efficacious in such groups as elder patients, smokers, patients with co-morbidities. Treatment of patients with co-morbidities and impacted health status will usually require increased resource utilization. Competitive recruitment may lead to imbalanced inclusion of the patients with potentially different co-morbidities and resource utilization (e.g. inclusion of elder patients in particular sites) and these patient-level differences may contribute to between-country variability in incremental resource use and effectiveness. Thus, inclusion of both patient-level and higher-level covariates will aid to understanding of heterogeneity. However, prediction of incremental results based on country-level covariates may not be feasible without applying of decision-analytical techniques (e.g. economic models).

Table 31: Selection Criteria of Covariates with Potential Effect on Incremental Health Economic Parameters

Level	Technical requirement	Considerations for selection of covariates	Potential covariates
Contextual/ higher-level (country, region, centre, investigator)	Recruited patients/study procedures are relevant/representative for the clinical population in particular setting with available macro-level parameters	▪ Clinical condition is chronic; ▪ New treatment introduces or substantially modify current clinical practice; ▪ New treatment may influence capacity utilization in provision of health-care services comparing to standard therapy; ▪ New treatment may change/influence incentives for health care providers and other relevant actors; ▪ New treatment require utilization/employment of cost-intensive techniques and/or high-skilled clinical personnel.	Clinical: Availability of prevention, screening, cessation programms; ratio of high-skilled to supportive personnel in hospitals; ratio of clinical personnel to number of beds. Socio-economic: %GDP dedicated to health care, number of hospital beds per capita, remuneration of clinical personnel, level of co-payment, location big city/country-side, university/excellence clinical site.

Level	Technical requirement	Considerations for selection of covariates	Potential covariates
Compositional/lower-level (individual patients)	Patient-level data available in study dataset	▪ Clinical condition is acute; ▪ Availability of clinical guidelines or treatment conventions suggesting different approaches (e.g. using of co-therapies) to different patients' sub-groups; ▪ Available evidence or indication for different efficacy and tolerability of new therapy in different sub-groups (elder patients, smokers, co-morbid, etc).	Clinical: Sex, age, co-morbidity, severity of condition, genetic predisposition. Socio-economic: Smoking, alcohol consumption, living environment, income, education, employment.

Comparative evaluation of the applied analytical generalizability methods

Table 32 summarizes advantages and limitations of selected generalizability methods for practical implementation of data analysis. Descriptive analysis of trial-wide and country-level results, while being easy to compute, have major limitations such as requirements of a large sample size and impossibility to account for covariates.

Qualitative and quantitative homogeneity tests are simple methods to access pronounced between-country heterogeneity. However, similar to descriptive statistics, the results can not be used as ultimate confirmation of the absence of treatment-by-country interaction. The calculation of test power is needed to evaluate robustness of test conclusions.

Hierarchical models offer flexible analytical framework using patient-level data with the option to include both patient- and country-level covariates, which is particularly useful to explain and account for observed between-country variability. The critics on the use of hierarchical models is that the exchangeability assumption can not hold when country-level covariates failed to explain between-country variability. Gelman and colleagues (2004) and Manca and Willan (2006) oppose to that suggesting, that despite between-country differences, it is feasible to consider the data drawn from a common distribution through conditional independence when building-in the covariates.

Table 32: Comparative Evaluation of Analytical Methods to Improve Generalizability

Method	Strengths	Limitations
Descriptive country-level analysis	- Direct estimation using country-specific parameters. - Applicable to large samples for sub-group analysis.	- High uncertainty about estimates of the small sub-groups. - Impossible formally to account for heterogeneity and covariates.

Method	Strengths	Limitations
Qualitative homogeneity test	Simple computation for detection of heterogeneity in effect's direction.	-Impossible to quantify sources of variation and account for covariates. - Low test power for economic parameters. - Absence of statistical significance should not mean absence of important economic differences.
Quantitative homogeneity test	Simple computation for detection of heterogeneity in effect's magnitude.	- Of limited use for country-level estimates since heterogeneity in magnitude can be expected. - Impossible formally to account for covariates.
Hierarchical models without covariates	- Estimation of country-level parameters taking into account hierarchical patient-level data. - Selections of appropriate distribution for costs or resource use parameters.	- Limited use if only few a priori different countries participated. - Assumption of "exchangeability" of patient-level data.
Hierarchical models with covariates	- Overcoming "exchangeability" assumption. - Exploring and explaining between-country differences with covariates.	- Possibly arbitrary selection of covariates. - High uncertainty of prediction of impact of covariates on incremental costs and effectiveness (and not on mean parameters only).

Source: compiled by the author.

What this case study adds

Methodology of this case study builds on previous research by Willan and Kowgier (2007), Nixon and Thompson (2005), Manca and colleagues (2005) and others. Our analysis contributes to the health economic literature on generalizability issue and methodology by following:

- Case study is first study in CAP indication, which explicitly addressed between-country heterogeneity and generalizability of the results of multinational patient-level health economic study.
- Case study demonstrated simultaneous application of several methods for assessment of between-country heterogeneity and improvement of generalizability in patient-level health economic study. Homogeneity test was found a complimentary method to hierarchical models to assess homogeneity and possibly "exchangeability" assumption for hierarchical modelling.
- Case study evaluated effect of selected country-level and patient-level covariates on mean and incremental costs and resource use in treatment of CAP indication.

In general, studying of between-country variability, analytical strategies to improve generalizability and use of appropriate methodology is still at the early development stage. Some findings of our case study may be valid for CAP indication only, however, analytical approach and major conclusions may to greater extend be relevant for other indications and medical treatments.

Limitations of the case study

Case study analysis has several limitations. The important limitation is a moderate sample size. This issue is important for the analysis of cost-effectiveness estimates based on data from 333 patients from 5 countries (hierarchical models 1-3). The issue with sample size is less relevant for the analysis of length of stay in the hospitals, as it was based on a larger sample of 622 patients from 10 countries. Nonetheless, health economic studies conducted along clinical trials are rarely powered to detect differences in economic parameters. Also, multinational studies are often conducted in many countries with the unbalanced and small sample size per country due to competitive

recruitment. Finally, depending on the indication, recruiting and conducting the trial with large sample size is often resource intensive. Thus, our case study could be considered as a representative dataset available for health economic analysis. In addition, cost-effectiveness analysis was conducted based on data from Western European countries. This fact supports assumptions about exchangeability of data in hierarchical models. The other limitation is the fact, that country-level covariates were not collected during the study conduct and analysis with covariates was not pre-specified in advance. As implication, we may not exclude spurious findings.

Implications and directions for future research

Studying variability and improving generalizability remain important future research topics in health economics. Both practical and theoretical aspects should be addressed. While methodology for analysis of multinational data made substantial progress, practical recommendations for the study design and conduct deserve more attention. In particular, guidelines for selection of countries, sites, number of patients per unit, selection and collection of potential covariates are needed. In addition, best practice and guidelines in application of costing methods are relevant to reduce variability caused by different methodological approaches to quantify resource use in different studies.

It will remain of interest to predict the cost-effectiveness outcomes beyond the study setting. This type of analysis would inevitably require evidence synthesis with help of meta-analysis and decision modelling. Exploring effects of country-level covariates on cost-effectiveness parameters might be helpful to make appropriate adjustment to the input data, while modelling cost-effectiveness for a setting of interest. For this purpose, theoretical and empirical basis for selection of appropriate covariates is needed taking into account both the nature of the condition (e.g. chronic or acute disease) and characteristics of the treatment options. In addition, the considerations about selection of appropriate covariates should be supported by the empirical health care system research and understanding which macroeconomic parameters are indicative for particular health care environment and could be associated with differences in mean and incremental cost-effectiveness estimates.

Another avenue for the research of statistical methods may explore usability of latent class modelling to uncover potential patient sub-groups or confirm/reject between- or within-country heterogeneity while simultaneously adjusting for covariates. It can be

also beneficial to explore impact of alternative distributional assumptions for dependent variables on cost-effectiveness estimates and covariates coefficients. There are a variety of possible extensions to the way covariates may be included in the model, such as non-linear transformations of variables, interactions among variables, and non-parametric regression effects, all of which can be incorporated into Bayesian hierarchical models.

Finally, assessment of between country-variability in different indications and study settings using various approaches and statistical techniques will contribute both to empirical and methodological knowledge in addressing generalizability issue in future research.

6 Conclusions

Health economic analysis will continue to inform and guide decision in health policy. Thus, generalizability of health economic evidence collected in multinational context, i.e. applicability of the study results to all participated countries and non-study settings, is important question for Public Health research and policy. In this work we explored the methods applicable for patient-level studies to improve generalizability of clinical and economic outcomes. The major findings can be summarized as following:

1) Common theoretical foundation rooted in Public Health and Economics is needed to guide health policy, study between-country variability and develop generalizability methods.

2) Homogeneity test and hierarchical models should be routinely used in the analysis to improve generalizability of the outcomes generated in patient-level studies.

3) Further practical, empirical and methodological research needed to explore requirements, applicability and limitations of the generalizability instruments in different datasets and contexts.

Numerous clinical and socio-economic factors are discussed in the literature and hypothesized to impact efficacy, resource use and costs of different technologies across the sites, countries and geographic areas. Research about the differences in health care delivery and outcomes is mainly driven by availability of empirical sources and lacks common theoretical foundations within Public Health and Economic realms. The approaches and methodology to explore between-setting variability and methods to improve generalizability of multinational health economic evaluations are yet to be developed and tested in different indications and datasets.

We applied homogeneity test and hierarchical models in secondary analysis of multinational study evaluating efficacy, resource use and costs of novel antibiotic treatment for acute respiratory condition. Homogeneity test supported assessment of significant between-country variability in clinical and economic parameters; hierarchical modelling allowed for calculation of trial-wide and country-level clinical and economic outcomes. Both methods complement standard health economic analysis

in assessment of statistical between-country heterogeneity and calculation of country-level and trial-wide clinical and economic endpoints with appropriate quantification of uncertainty.

Decision about homogeneity of the targeted outcomes may not exclusively rely on the results of homogeneity test, in particular, when comparing results across the countries with small samples. Hierarchical models support estimation of trial-wide and setting-specific results. Validity of "exchangeability" assumption for application of hierarchical modelling, i.e. no a priori reasons to consider the settings explicitly different, is vital and should be addressed in the analysis. In addition, the method allows for adjustment of differences in patient-level baseline variables. Predictive ability of country-level covariates to support generalizability of incremental costs, resource use and effectiveness to the countries not participated in the study merits further investigation. The differences in socio-economic context across the countries, such as proportion of GDP dedicated to health care, capacity of health facilities and other factors, may or may not have effect on incremental clinical and economic parameters. Reliable assessment of the association between country-level covariates and incremental effects requires large sample size from several countries.

Generalizability subject should be routinely addressed at the design and analysis phases of multinational health economic studies. Future research should focus on both methodological development, as well as application of available methods in different data sets and clinical indications. The collected empirical data will provide further inside on requirements and limitations of used instruments.

Analysis of multinational evidence has several implication for health policy. Assessment of the sources of between-country variability will contribute to better understanding of the potential reasons of (in)efficient health production and thus contribute to formulation of informed health policy.

References

Alban, A., Gyldmark, M., Pedersen, A. V., & Søgaard, J. (1997). The Danish approach to standards for economic evaluation methodologies. PharmacoEconomics, 12, 627-636.

Alker, H. S. (1969). A typology of ecological fallacies. In M. Dogan & S. Rokkan (Eds.), Quantitative ecological analysis. MA, USA: MIT Press.

AMCP. (2005). The AMCP format for formulary submissions (US). Retrieved Mar, 2009, from http://www.amcp.org/home.

American Thoracic Society. (1993). Guidelines for the initial management of adults with community acquired pneumonia: Diagnosis, assessment of severity, and initial antimicrobial therapy. The American Review of Respiratory Diseases, 148, 1418-1426.

Anand, K., Pandav, C., & Nath, L. M. (1999). Impact of HIV/AIDS on the national economy of India. Health Policy, 47, 195-205.

Anderson, F., Reinhardt, U., Hussey, P., & Petrosyan, V. (2003). It's the prices, stupid: why the United States is so different from other countries. Health Affairs, 22(3), 89-105.

Antonovsky, A. (1979). Health, stress and coping. San Francisco: Jossey-Bass.

Antonovsky, A. (1987). Unraveling the mystery of health. How people manage stress and stay well. San Francisco: Jossey-Bass.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. American Economic Review, 53, 941-973.

Ashton, T. (1998). Contracting for health services in New Zealand: a transaction cost approach. Social Science and Medicine, 46(3), 357-367.

Babazono, A., & Hillman, A. (1994). A comparison of international health outcomes and health care spending. International Journal of Technology Assessment in Health Care, 10(03), 376-381.

Baltussen, R., Leidl, R., & Ament, A. (1999). Real world designs in economic evaluation. Bridging the gap between clinical research and policy-making. Pharmacoeconomics, 16(5), 449-458.

Bank of Canada. (2008). International exchange rates. Retrieved Sep, 2008, from http://www.bankofcanada.ca/en/rates/exchform.html.

Banta, D. (2003). Considerations in defining evidence for Public Health. The European Advisory Committee on Health Research. World Health Organization, Regional Office for Europe. International Journal of Technology Assessment in Health Care, 19(3), 559-572.

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

Bartlett, J. G., & Mundy, L. M. (1995). Community-acquired pneumonia. New English Journal of Medicine, 333(24), 1618-1624.

Bartlett, W. (1991). Quasi-markets and contracts: a markets and hierarchies perspective on NHS reform. Public Money and Management, 11(3), 53-61.

Battleman, D., Callahan, M., & Thaler, H. (2002). Rapid antibiotic delivery and appropriate antibiotic selection reduce length of hospital stay of patients with community-acquired pneumonia: Link between quality of care and resource utilization. Archives of Internal Medicine, 162(6), 682-688.

Beard, S. M., Coleman, R., Radford, J., & Tidy, J. (1997). The use of cisplatin and paclitaxel as a first line treatment in ovarian cancer: Universities of Leicester, Nottingham and Sheffield.

Becker, G. S. (1964). Human Capital: A theoretical and empirical Analysis, with special reference to education. Chicago: University of Chicago Press.

Behmane, D., Lambot, K., Irs, A., & Seikunas, N. (2002). Baltic guideline for economic evaluation of pharmaceuticals (pharmacoeconomic analysis) [Electronic Version]. Retrieved Dec, 2009 from http://www.zca.gov.lv/english/guidelines.html.

Birch, S., & Donaldson, C. (2003). Valuing the benefits and costs of health care programmes: Where's the 'extra' in extra-welfarism? Social Science & Medicine, 56(5), 1121-1133.

Birch, S., & Gafni, A. (2003). Economics and the evaluation of health care programmes: Generalisability of methods and implications for generalisability of results. Health Policy, 64(2), 207-219.

Bombardier, C., & Maetzel, A. (1999). Pharmacoeconomic evaluation of new treatments: Efficacy versus effectiveness studies? Annals of the Rheumatic Diseases, 58(90001), 82-85.

Boulenger, S., Nixon, J., Drummond, M., Ulmann, P., Rice, S., & de Pouvourville, G. (2005). Can economic evaluations be made more transferable? European Journal of Health Economics, 6(4), 334-346.

Brennan, A., & Akehurst, R. (2000). Modelling in health economic evaluation. What is its place? What is its value? Pharmacoeconomics, 17(5), 445-459.

Briggs, A., & Gray, A. (1999). Methods in health service research: Handling uncertainty in economic evaluations of healthcare interventions. BMJ, 319(7210), 635-638.

British National Formulary. (2001). Joint Formulary Committee. British Medical Association and Royal Pharmaceutical Society of Great Britain. London, UK. Retrieved from www.bnf.org

British Thoracic Society. (1993). Guidelines for the management of community-acquired pneumonia in adults admitted to hospital. British Journal of Hospital Medicine, 49(5), 346-350.

Britton, M. (1990). Geographical variation in mortality since 1920 for selected causes. In M. Britton (Ed.), Mortality and Geography: A Review in the mid-1980's for England and Wales. London: HMSO.

Broome, J. (1991). Weighting goods: equality, uncertainty and time. Oxford: Oxford University Press.

Brouwer, W., Exel, J. V., Baal, P. V., & Polder, J. (2006). Economics and public health: Engaged to be happily married. The European Journal of Public Health, September(17), 122-123.

Brown, L. C., Epstein, D., Manca, A., Beard, J. D., Powell, J. T., & Greenhalgh, R. M. (2004). The UK endovascular aneurysm repair (EVAR) trials: Design, methodology and progress. European Journal of Vascular and Endovascular Surgery, 27(4), 372-381.

Bruning, E., & Register, C. (1989). Technical efficiency within hospitals: Do profit incentives matter? Applied Economics, 21(9), 1217-1234.

Bryan, S., & Brown, J. (1998). Extrapolation of cost-effectiveness information to local settings. Journal of Health Services Research&Policy, Apr 3(2), 108-112.

Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Newbury Park, CA: Sage Publications.

BSP study report. (2000). BSP study report: BAY 12-8039/IMPACT-200036. Study SAS database. BSP Internal document.

Buchanan, J. (1963). The economics of earmarked taxes. Journal of Political Economy, 71, 457.

Call to Action: Health Reform (US) [Electronic (2008). Version]. White Paper. Retrieved Nov, 2009 from http://finance.senate.gov/healthreform2009/finalwhitepaper.pdf.

Canadian Agency for Drugs and Technologies in Health. (2006). Guidelines for the economic evaluation of health technologies: Canada. 3rd edition. Retrieved Jul, 2009, from http://www.cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf.

Capri, S., Ceci, A., Terranova, S., Merlo, F., & Mantovani, L. (2001). Guidelines for economic evaluations in Italy: Recommendations from the italian group of pharmacoeconomic studies. Drug Information Journal, 35, 189-201.

Carey, K. (2000). A multilevel modelling approach to analysis of patient costs under managed care. Health Economics, 9(5), 435-446.

Caro, J., Klittich, W., McGuire, A., Ford, I., Pettitt, D., Norrie, J., et al. (1999). International economic analysis of primary prevention of cardiovascular disease with pravastatin in WOSCOPS. European Heart Journal, 20(4), 263-268.

Caro, J. J., Huybrechts, K. F., De Backer, G., De Bacquer, D., & Closon, M. C. (2000). Are the WOSCOPS clinical and economic findings generalizable to other populations? A case study for Belgium. The WOSCOPS Economic Analysis Group. West of Scotland Coronary Prevention Study. Acta Cardiologica, 55(4), 239-246.

Carr-Hill, R. A., Rice, N., & Roland, M. (1996). Socioeconomic determinants of rates of consultation in general practice based on fourth national morbidity survey of general practices. BMJ, 312(7037), 1008-1012.

Clarkson, K. W. (1972). Some implications of property rights in hospital management. Journal of Law and Economics, 15, 363-384.

Claxton, K. (1999). The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. Journal of Health Economics, 18(3), 341-364.

Claxton, K., Sculpher, M., & Drummond, M. (2002). A rational framework for decision making by the National Institute For Clinical Excellence (NICE). The Lancet, 360(9334), 711-715.

Coase, R. (1937a). Earnings of skilled workers in a manufacturing enterprise. Economica, 4, 115-116.

Coase, R. (1937b). The nature of the firm. Economica, 4, 386-405.

Collège des Économistes de la Santé. (2004). French guidelines for the economic evaluation of health care technologies [Electronic Version]. Retrieved Jun, 2009 from http://www.ces-asso.org.

College voor zorgverzekeringen Diemen. (2006). Guidelines for pharmacoeconomic research, updated version. [Electronic Version]. Retrieved Jun, 2009 from http://www.ispor.org/peguidelines/source/HTAGuidelinesNLupdated2006.pdf.

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

Cook, J., Drummond, M., & Heyse, J. (2004). Economic endpoints in clinical trials. Statistical Methods in Medical Research, 13(2), 157-176.

Coughlin, C. M., Nelson, M., Merchant, S., & Gondek, K. (2003). Costs of broad-spectrum antibiotic use for acute sinusitis, chronic bronchitis, and pneumonia in a managed care population. Managed Care Interface, 16(6), 34-40, 55.

Coyle, D. (2003). Analyzing uncertainty and variability in economic evaluation: an application to the treatment of osteoporosis. Meeting of the International Health Economics Association, 4, 282-283.

Coyle, D., & Drummond, M. F. (2001). Analyzing differences in the costs of treatment across centers within economic evaluations. International Journal of Technology Assessment of Health Care, 17(2), 155-163.

Coyle, D., & Lee, K. M. (1998). The problem of protocol driven costs in pharmacoeconomic analysis. Pharmacoeconomics, Oct 14(4), 357-363.

Culyer, A. J. (1990). Commodities, characteristics of commodities, characteristics of people, utilities and the quality of life. In S. Baldwin, C. Godfrey & C. Propper (Eds.), Quality of life: perspectives and policies (pp. 9-27). London: Routledge.

Culyer, A. J. (1998). How ought health economists to treat value judgements in their analyses? In M. Barer, T. Getzen & G. Stoddart (Eds.), Health, health care and health economics (pp. 363-371). Chichester: Wiley.

Culyer, A. J., & Evans, R. G. (1996). Mark Pauly on welfare economics: Normative rabbits from positive hats. Journal of Health Economics, 15(2), 243-251.

Davis, P., & Gribben, B. (1995). Rational prescribing and interpractitioner variation. A multilevel approach. International Journal of Technology Assessment in Health Care, 11(2), 428-442.

DiPrete, T., & Forristal, J. (1994). Multilevel models: Methods and substance. Annual Review of Sociology, 20, 331-357.

Documed. (2001). Compendium Suisse des medicaments. Documed AG, Basel, Switzerland. Retrieved Mar, 2008 from http://www.kompendium.ch/Search.aspx.

Dolan, P., & Edlin, R. (2002). Is it really possible to build a bridge between cost-benefit analysis and cost-effectiveness analysis? Journal of Health Economics, 21, 827-843.

Draper, D. (1995). Inference and hierarchical modeling in the social sciences. Journal of Educational and Behavioral Statistics, 20(2), 115-147.

Dresser, L. D., Niederman, M. S., & Paladino, J. A. (2001). Cost-effectiveness of gatifloxacin vs ceftriaxone with a macrolide for the treatment of community-acquired pneumonia. Chest, 119(5), 1439-1448.

Drummond, M. (1991). Output measurement for resource-allocation decisions in health care. In A. McGuire, P. Fenn & K. Mayhew (Eds.), Providing health care. The economics of alternative systems of finance and delivery. Oxford: Oxford University Press.

Drummond, M., Barbieri, M., Cook, J., Glick, H. A., Lis, J., Malik, F., et al. (2009). Transferability of economic evaluations across jurisdictions: ISPOR good research practices task force report. Value Health, 12(4), 409-418.

Drummond, M., Manca, A., & Sculpher, M. (2005). Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies. International Journal of Technology Assessment in Health Care, 21(2), 165-171.

Drummond, M., & Pang, F. (2001). Transferability of results of economic evaluations. In M. Drummond & A. McGuire (Eds.), Economic Evaluation in Health Care: Merging Theory with Practice Oxford: Oxford University Press.

Drummond, M. F., Becker, D. L., Hux, M., Chancellor, J. V. M., Duprat-Lomon, I., Kubin, R., et al. (2003). An economic evaluation of sequential IV/po moxifloxacin therapy compared to IV/po co-amoxiclav with or without clarithromycin in the treatment of community-acquired pneumonia. Chest, 124(2), 526-535.

Drummond, M. F., O'Brien, B. J., Stoddart, G., & Torrance, G. (1997). Methods for the economic evaluation of health care programmes (2nd ed.). Oxford: Oxford University Press.

Duncan, C., Jones, K., & Moon, G. (1996). Health-related behaviour in context: A multilevel modelling approach. Social Science & Medicine, 42(6), 817-830.

Duncan, C., Jones, K., & Moon, G. (1998). Context, composition and heterogeneity: Using multilevel models in health research. Social Science&Medicine, 46(1), 97-117.

EARSS. (2006). EARSS Annual Report [Electronic Version]. Retrieved Jul, 2008 from www.rivm.nl/earss.

EMEA. (2002). Good Clinical Practice for Trials on Medicinal Products in the European Community (CPMP). Retrieved Nov, 2009, from http://www.ema.europa.eu/pdfs/human/ich/013595en.pdf.

Evans, C. J., & Crawford, B. (2000). Data collection methods in prospective economic evaluations: How accurate are the results? Value in Health, 3(4), 277-286.

Ewig, S., Reinert, R., Wiedemann, B., Rodloff, A., Trautmann, M., Pletz, M., et al. (2002). Antibiotika-Resistenz bei Erregern ambulant erworbener Atemwegsinfektionen. Chemotherapie, 11, 12-26.

Fenwick, E., O'Brien, B. J., & Briggs, A. H. (2004). Cost-effectiveness acceptability curves – facts, fallacies and frequently asked questions. Health Economics, 13, 405-415.

File, T. M. (2003). Community-acquired pneumonia. Lancet, 362(9400), 1991-2001.

File, T. M., Larsen, L. S., & Fogarty, C. M. (2001). Safety and efficacy of sequential (i.v. to po) moxifloxacin for the treatment of community-acquired pneumonia in hospitalized patients. Today's Therapeutic Trends, 19, 251-270.

Finch, R., Schurmann, D., Collins, O., Kubin, R., McGivern, J., Bobbaers, H., et al. (2002). Randomized controlled trial of sequential intravenous (i.v.) and oral moxifloxacin compared with sequential i.v. and oral co-amoxiclav with or without clarithromycin in patients with community-acquired pneumonia requiring initial parenteral treatment. Antimicrobial Agents and Chemotherapy, 46(6), 1746-1754.

Fine, M. J., Auble, T. E., Yealy, D. M., Hanusa, B. H., Weissfeld, L. A., Singer, D. E., et al. (1997). A prediction rule to identify low-risk patients with community-acquired pneumonia. New England Journal of Medicine, 336(4), 243-250.

Fine, M. J., Pratt, H. M., Obrosky, D. S., Lave, J. R., McIntosh, L. J., Singer, D. E., et al. (2000). Relation between length of hospital stay and costs of care for patients

with community-acquired pneumonia. The American Journal of Medicine, 109(5), 378-385.

Fine, M. J., Singer, D. E., Hanusa, B. H., Lave, J. R., & Kapoor, W. N. (1993). Validation of a pneumonia prognostic index using the MedisGroups Comparative Hospital Database. The American Journal of Medicine, 94(2), 153-159.

Fine, M. J., Smith, D. N., & Singer, D. E. (1990). Hospitalization decision in patients with community-acquired pneumonia: a prospective cohort study. The American Journal of Medicine, 89(6), 713-721.

Folland, S., Goodman, A. C., & Stano, M. (2006). The economics of health and health care (5 Ed.). NJ: Prentice Hall.

Forsberg, E., Axelsson, R., & Arnetz, B. (2001). Financial incentives in health care: the impact of performance-based reimbursement. Health Policy, 58(3), 243-262.

Frech, H. E. (1976). The Property Rights Theory of the firm: empirical results from a natural experiment. Journal of Political Economy, 84, 143-152.

Frech, H. E. (1985). The Property Rights Theory of the firm: some evidence from the U.S. nursing home industry. Journal of Institutional and Theoretical Economics, 14, 146-166.

Friedman, B., & Pauly, M. (1981). Cost functions for a service firm with variable quality and a stochastic demand: the case of hospitals. Review of Economics and Statistics, 63, 620-624.

Fryback, G. D., Chinnis, J. O., & Ulvila, J. W. (2001). Bayesian cost-effectiveness analysis. An example using the GUSTO trial. International Journal of Technology Assessment in Health Care, 17(1), 83-97.

Gail, M., & Simon, R. (1985). Testing for qualitative interactions between treatment effects and patient subsets. Biometrics, 41, 361-372.

Garber, A. M., & Phelps, C. E. (1997). Economic foundations of cost-effectiveness analysis. Journal of Health Economics, 16, 1-31.

Garcia-Ordonez, M., Garcia-Jimenez, J., Paez, F., Álvarez, F., Poyato, B., Franquelo, J., et al. (2001). Clinical aspects and prognostic factors in elderly patients

hospitalised for community acquired pneumonia. European Journal of Clinical Microbiology & Infectious Diseases, 20, 14-19.

Garson, G. D. (2009). Statnotes: Topics in Multivariate Analysis. Multilevel Modeling. Retrieved Dec, 2009, from http://faculty.chass.ncsu.edu/garson/PA765/statnote.htm.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). Bayesian Data Analysis. New York: Chapman & Hall/CRC.

Ghatnekar, O., Persson, U., Willis, M., & Odegaard, K. (2001). Cost effectiveness of Becaplermin in the treatment of diabetic foot ulcers in four European countries. Pharmacoeconomics, 19(7), 767-778.

Gilks, W., Richardson, S., & Spiegelhalter, D. (1996). Markov chain Monte Carlo in practice. London: Chapman and Hall.

Glick, H., Orzol, S. M., Tooley, D., Polsky, D., & Mauskopf, J. O. (2003). Design and analysis of unit cost estimation studies: How many hospital diagnoses? How many countries? Health Economics, 12(7), 517-527.

GOA. (2000). Gebuerenordnung fuer Aerzte (GOA) (Juli 2001, Stand: 1.1 ed.). Germany: Deutscher Aerzte-Verlag.

Goeree, R., Burke, N., O'Reilly, D., Manca, A., Blackhouse, G., & Tarride, J. (2007). Transferability of economic evaluations: approaches and factors to consider when using results from one geographic area for another. Current Medical Research and Opinion, 23(4), 671-682(612).

Goeree, R., Gafni, A., Hannah, M., Myhr, T., & Blackhouse, G. (1999). Hospital selection for unit cost estimates in multicentre economic evaluations: Does the choice of hospitals make a difference? Pharmacoeconomics, 15(6), 561-572.

Goldstein, H. (1995). Multilevel statistical models. London: Edward Arnold.

Goldstein, H. (1999). Multilevel statistical models. Retrieved Feb, 2010, from http://www.soziologie.uni-halle.de/langer/multilevel/books/goldstein.pdf.

Goodacre, S., & Dixon, S. (2005). Is a chest pain observation unit likely to be cost-effective at my hospital? Extrapolation of data from a randomized controlled trial. Emergency Medicine Journal, 22, 418-422.

Goodman, C. (2004). Introduction to health technology assessment [Electronic Version]. HTA 101. Retrieved Jan 2010 from http://www.nlm.nih.gov/nichsr/hta101/hta101.pdf.

Goodman, C. A., & Mills, A. J. (1999). The evidence base on the cost-effectiveness of malaria control measures in Africa. Health Policy Plan, 14, 301-312.

Greiner, W., Schöffski, O., & Graf v. d. Schulenburg, J. M. (2000). The transferability of international economic health-economic results to national study questions. Health Economics in Prevention and Care (HEPAC), 1(2), 94-102.

Grieve, R., Nixon, R., Thompson, S., & Normand, C. (2005). Using multilevelmodels for assessing the variability of multinational resource use and cost data. Health Economics, 14, 185-196.

Grieve, R., Thompson, S., Nixon, R., & Cairns, J. (2007). Multilevel models for estimating incremental net benefits in multinational studies. Health Economics, 16, 815-826.

Guest, J., & Morris, A. (1997). Community-acquired pneumonia: the annual cost to the National Health Service in the UK. European Respiratory Journal, 10(7), 1530-1534.

Hall, J., Gafni, A., & Birch, S. (2006). Health economics critiques of welfarism and their compatibility with Sen's capabilities approach. CHERE Working Paper, 16(September).

Hand, W. L., Boozer, R. M., & King-Thompson, N. L. (1985). Antibiotic uptake by alveolar macrophages of smokers. Antimicrobial Agents and Chemotherapy, Jan, 42-45.

Hart, O. (1995). Firms, contracts and financial structure. Oxford: Oxford University Press.

Hart, O. (2003). Incomplete contracts and public ownership: Remarks and an application to public–private partnerships. Economic Journal, 113(468), C69-76.

Heusser, P. (2001). Kriterien zur Beurteilung des Nutzens von komplementärmedizinischen Methoden. Forsch Komplementärmed Klass Naturheilkd, 8, 14-23.

Hirschman, A. (1970). Exit, voice, and loyalty: responses to decline in firms, organizations, and states. Cambridge, MA: Harvard University Press.

Hox, J. J., & Kreft, G. G. (1994). Multilelevel analysis methods. Sociological Methods and Research, 22, 283-289.

Hughes, D., Griffiths, L., & McHale, J. (1997). Do quasi-markets evolve? Institutional analysis and the NHS. Cambridge Journal of Economics, 21, 259-276.

Hurley, J. (1998). Welfarism, extra-welfarism, and evaluative economic analysis in the health sector. In M. Barer, T. Getzen & G. Stoddart (Eds.), Health, Health Care and Health Economics: Perspectives on Distribution (pp. 373-396). Toronto: John Wiley and Sons.

Hurley, J. (2000). An overview of the Normative Economics of the health sector. In J. A. Culyer & J. Newhouse (Eds.), Handbook of Health Economics (Vol. 1, pp. 55-118). Elsevier Science.

Hutton, J., McGrath, C., Frybourg, J. M., Tremblay, M., Bramley-Harker, E., & Henshall, C. (2006). Framework for describing and classifying decision-making systems using technology assessment to determine the reimbursement of health technologies (fourth hurdle systems). International Journal of Technology Assessment in Health Care, 22(1), 10-18.

IHS Global Insight. (2008). World Market Healthcare. Retrieved Mar 2008, from http://www.ihsglobalinsight.com/ProductsServices/ProductDetail794.htm

IMS Health. (2002). Market share data. IMS data online. Retrieved Mar 2008, from http://www.imshealth.org.

Infarmed. (1998). Guidelines For Economic Drug Evaluation Studies. Ministery of Health, Lisbon, Portugal.

IQWiG. (2008). Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Allgemeine Methoden. Version 3. Retrieved Mar, 2010, from http://www.iqwig.de/index.de.html.

IQWiG. (2009). Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Entwurf einer Methodik für die Bewertung von Verhältnissen zwischen Nutzen und Kosten im System der deutschen gesetzlichen Krankenversicherung. Version 2. Retrieved Mar, 2010, from http://www.iqwig.de/index.de.html.

IPHA/HSE. (2000). Irish Healthcare Technology Assessment Guidelines (Irland). Retrieved Mar 2009, from http://www.ncpe.ie/document.php?cid=26&sid=65&docid=62.

ISPOR. (2010). Pharmacoeconomic guidelines around the world. Retrieved Jan 2010, from http://www.ispor.org/peguidelines/index.asp.

Janssens, J. P. (2005). Pneumonia in the elderly (geriatric) population. Current Opinion in Pulmonary Medicine, 11(3), 226-230.

Johannesson, M., & Karlsson, G. (1997). The friction cost method: a comment. Journal of Health Economics, 16, 249-256.

Jonsson, B., & Weinstein, M. C. (1997). Economic evaluation alongside multinational clinical trials. International Journal of Technology Assessment in Health Care, 13, 49-58.

Jonsson, E. (2002). Development of Health Technology Assessment in Europe. International Journal of Technology Assessment in Health Care, 18(2), 171-183.

Jonsson, E., & Banta, D. (1999). Management of health technologies: an international view. BMJ, 319, 1293.

Joskow, P. L. (2008). Introduction to New Institutional Economics: A report card. In E. Brousseau & J. M. Glachant (Eds.), New Institutional Economics: A Guidebook (pp. 1-20). New York: Cambridge University Press.

Katz, E., Larsen, L. S., Fogarty, C. M., Hamed, K., Song, J., & Choudhri, S. (2004). Safety and efficacy of sequential i.v. to p.o. moxifloxacin versus conventional combination therapies for the treatment of community-acquired pneumonia in patients requiring initial i.v. therapy. Journal of Emergency Medicine, 27(4), 395-405.

KCE. (2008). Pharmaco-Economic Guideline. KCE reports vol. 28. Belgian Health Care Knowledge Centre. Retrieved Mar 2009, from http://www.kce.fgov.be/index_en.aspx?SGREF=5211.

Keen, J., & Ferguson, B. (1996). Transaction costs, externalities and information technology in health care. Health Economics, 5, 25-36.

Kim, J., & Mahoney, J. T. (2005). Property Rights Theory, Transaction Costs Theory, and Agency Theory: An Organizational Economics approach to strategic management. Managerial and Decision Economics, 26, 223-242.

Kinosian, B., & Glick, H. (1998). Clinical trials and external validity. Value in Health, 1(3), 181-185.

Kobelt, G., Jonsson, L., & Fredrikson, S. (2003). Cost-utility of interferon beta1b in the treatment of patients with active relapsing-remitting or secondary progressive multiple sclerosis. European Journal of Health Economics, 4(1), 50-59.

Koopmanschap, M. A., Touw, K., & Rutten, F. H. (2001). Analysis of costs and cost-effectiveness in multinational trials. Health Policy, 58(2), 175-186.

Kühn, H., & Rosenbrock, R. (1994). Präventionspolitik und Gesundheitswissenschaften. Eine Problemskizze In R. Rosenbrock, H. Kühn & B. Köhler (Eds.), Präventionspolitik. Gesellschaftliche Strategien der Gesundheitssicherung (pp. 29-53). Berlin: Edition sigma.

Kulp, W., & Greiner, W. (2006). Gesundheitsökonomie und HTA. Bundesgesundheitsblatt - Gesundheitsforschung – Gesundheitsschutz, 3(49), 257-263.

Lave, J. R., Lin, C. J., Hughes-Cromwick, P., & Fine, M. J. (1999). The cost of treating patients with community-acquired pneumonia. Seminars in Respiratory and Critical Care Medicine, 20, 189.

Leyland, A. H., & Boddy, F. A. (1997). Measuring performance in hospital care: Length of stay in gynaecology. The European Journal of Public Health, 7(2), 136-143.

Li, S. C. (2007). How to improve pharmacoeconomic data generalisability between different countries. Expert Opinion on Pharmacotherapy, 8(10), 1409-1413.

Lode, H., Grossman, C., Choudhri, S., Haverstock, D., McGivern, J., Herman-Gnjidic, Z., et al. (2003). Sequential IV/PO moxifloxacin treatment of patients with severe community-acquired pneumonia. Respiratory Medicine, 97(10), 1134-1142.

Longford, N. T. (1993). Random coefficient models. New York: Oxford University Press.

Low, D. E. (2004). Quinolone resistance among pneumococci: therapeutic and diagnostic implications. Clinical Infectious Diseases, 38, Suppl 4, 357-362.

Luetzenkirchen, A. (2000). Sozialer Wandel und Veränderung der Gesundheitslage ostdeutscher Frauen. In Laaser & Ulrich (Eds.), Gesundheit und soziale Benachteiligung. Informationssysteme - Bedarfsanalysen - Interventionen: Hans Jacobs.

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

Manca, A., Rice, N., Sculpher, M. J., & Briggs, A. H. (2005). Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models. Health Economics, 14(5), 471-485.

Manca, A., & 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.

Mandell, L. A., Iannini, P. B., & Tillotson, G. S. (2004). Respiratory fluoroquinolones: differences in the details. Clinical Infectious Diseases, 38(9), 1331-1332; author reply 1332-1333.

Mankiw, G. (2008). Principles of Economics (5 ed.). Mason, USA: South-Western Cengage Learnings.

Mannion, R., & Small, N. (1999). Postmodern health economics. Health Care Analysis, 7, 255-272.

Marini, G., & Street, A. (2007). A Transaction Costs analysis of changing contractual relations in the English NHS. Health Policy, 83, 17-26.

Mark, W. (2008). Epidemiology of adult hospitalized community-acquired pneumonia. In R. M. Antoni Torres (Ed.), Community-Acquired Pneumonia (pp. 5-19). England: John Wiley & Sons.

Marrie, T. J. (1998). Community-acquired pneumonia: epidemiology, etiology, treatment. Infectious Disease Clinics of North America,12, 723-740.

Mason, J., & Drummond, M. (1997). The generalisability of pharmacoeconomic studies. Pharmacoeconomics, 11, 503-514.

Mason, J. M., & Mason, A. R. (2006). The generalisability of pharmacoeconomic studies: Issues and challenges ahead. Pharmacoeconomics, 24(10), 937-945.

McCormick, D., Fine, M. J., Coley, C. M., Marrie, T. J., Lave, J. R., Obrosky, D. S., et al. (1999). Variation in length of hospital stay in patients with community-acquired pneumonia: are shorter stays associated with worse medical outcomes? The American Journal of Medicine, 107(1), 5-12.

McPake, B., & Normand, C. E. M. (2002). Health economics: An international perspective. London and New York: Routlege.

Milgrom, P., & Roberts, J. (1992). Economics of organization and management. Prentice-Hall, NJ: Englewood Cliffs.

Ministry of Social Affairs and Health. (2003). Guideline for preparing a health economic evaluation. Retrieved Mar, 2009, from http://www.ispor.org/peguidelines/countrydet.asp?c=7&t=1.

Monge, V., San-Martin, V. M., & Gonzalez, A. (2001). The burden of community-acquired pneumonia in Spain. European Journal of Public Health, 11, 362-364.

Mooney, G. (1998). "Communitarian claims" as an ethical basis for allocating health care resources. Social Science & Medicine, 47, 1171-1180.

Mooney, G. (2005). Communitarian claims and community capabilities: furthering priority setting? Social Science & Medicine, 60(2), 247-255.

Mooney, G., & Jan, S. (1997). Vertical equity: Weighting outcomes? or establishing procedures? Social Science & Medicine, 39, 79-87.

Mooney, G., & Russell, E. (2003). Equity in health care: The need for a new economics paradigm? In A. Scott, A. Maynard & R. Elliott (Eds.), Advances in Health Economics (pp. 205-221). Chichester: John Wiley & Sons Ltd.

Murray, C., Evans, D. B., Acharya, A., & Baltussen, R. (2000). Development of WHO guidelines on generalized cost-effectiveness analysis. Health Economics, 9(3), 235-251.

Musgrove, P., & Fox–Rushby, J. (2006). Cost–effectiveness analysis for priority setting. In Disease control priorities in developing countries (2nd ed., pp. 271-286). New York: Oxford University Press.

NGAP. (2001). Nomenclature Generale des Actes professionels (NGAP). L'Union des Caisses Nationales de Securite (UCANSS). Paris, France.

NHS Drug Tariff. (2001). National Health Service England and Wales. Retrieved Mar 2008, from http://www.ppa.org.uk/edt/March_2010/mindex.htm.

NHS Reference Costs. (2001/2002). Deparetment of Health (DH). NHS. Retrieved Mar 2008, from http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4069646.

NICE. (2008). Guide to the methods of technology appraisal. National Institute for Clinical Excellence, NHS. June. Retrieved Mar, 2009, from http://www.nice.org.uk .

NICE. (2009). Guide to the single technology appraisal process. National Institute for Clinical Excellence, NHS. Retrieved Jan, 2010, from http://www.nice.org.uk.

Niederman, M. S., Mandell, L. A., Anzueto, A., Bass, J. B., Broughton, W. A., Campbell, G. D., et al. (2001). Guidelines for the management of adults with community-acquired pneumonia. Diagnosis, assessment of severity, antimicrobial therapy, and prevention. American Thoracic Society. American Journal of Respiratory and Critical Care Medicine, 163(7), 1730-1754.

Niederman, M. S., McCombs, J. S., Unger, A. N., Kumar, A., & Popovian, R. (1998). The cost of treating community-acquired pneumonia. Clinical Therapy, 20, 820-837.

Nixon, R. M., & Thompson, S. G. (2005). Methods for incorporating covariate adjustment, subgroup analysis and betweencentre differences into cost-effectiveness evaluations. Health Economics, 14, 1217-1229.

North, D. C. (1990). Institutions, institutional change and economic performance. Cambridge: Cambridge University Press.

North, D. C. (2005). Understanding the process of economic change. Princeton: Princeton University Press.

Norwegian Medicine Agency. (2005). Norwegian guidelines for pharmacoeconomic analysis in connection with applications for reimbursement. Retrieved Mar, 2009, from http://www.ispor.org/peguidelines/countrydet.asp?c=15&t=1.

O'Brien, B. J. (1997). A tale of two (or more) cities: geographic transferability of pharmacoeconomic data. American Journal of Managed Care, 3(Suppl 5), 33-39.

OECD. (2005). Health at a Glance: OECD Indicators 2005. Source OECD Social Issues/Migration/Health(November), i-175(176).

OECD. (2008). OECD Health Data 2008. Statistics and indicators for 30 countries. Paris: OECD Publishing. Retrieved Jan 2009, from www.oecd.org/health/healthdata.

OECD. (2009). OECD Statistics. Price indices and purchasing power parity data Retrieved Jan 2010, from http://stats.oecd.org/Index.aspx?querytype=view&queryname=222.

Or, Z. (2000). Determinants of health outcomes in industrialised countries: a pooled cross-country, time-series analysis. OECD Economic Studies, 30(1), 53-77.

Orlewska, E., & Mierzejewski, P. (2004). Proposal of Polish Guidelines for conducting financial analysis and their comparison to existing guidance on budget impact in other countries (Poland). Value in Health, 7(1), 1-10.

Palmer, C. S., Zhan, C., Elixhauser, A., Halpern, M. T., Rance, L., Feagan, B. G., et al. (2000). Economic assessment of the community-acquired pneumonia intervention trial employing levofloxacin. Clinical Therapy, 22(2), 250-264.

Palmer, S., Sculpher, M., Philips, Z., Robinson, M., Ginelly, L., Bakhai, A., et al. (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, 229-240.

Pan, G., & Wolfe, D. A. (1997). Test for qualitative interaction of clinical significance. Statistics in Medicine, 16, 1645-1652.

Paterson, L., & Goldstein, H. (1992). New statistical methods for analyzing social structures: an introduction to multilevel models. British Educational Research Journal, 17, 387-393.

Peltzman, S. (1976). Toward a more general theory of regulation. Journal of Law & Economics, University of Chicago Press, 19(2), 211-240.

Personal Social Services Research Unit (PSSRU). (2001). Retrieved Mar, 2009, from http://www.lse.ac.uk/collections/PSSRU/.

PHARMAC. (2007). Prescription for pharmacoeconomic analysis. Methods for cost-utility analysis. Retrieved May, 2009, from http://www.pharmac.govt.nz/EconomicAnalysis/pharmacoeconomics.

Physioswiss. (2001). Tarifvertrag. Schweizerischen Physiotherapeutenverband (SPV). Retrieved Mar 2008, from http://www.physioswiss.ch/index.cfm?navid=207

Physiotherapy offices North Rhein-Westphalia. (2001). Services, tariffs and remuniration. Deutscher Verband für Physiotherapie-Zentralverband der Physiotherapeuten / Krankengymnasten (ZVK) e.V., Landesverband NRW. Retrieved from http://www.zvknrw.de.

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

PKV. (1999/2000). Die private Krankenversicherung. Zahlenbericht 1999/2000. Verband der Private Krankenversicherung (PKV). Retrieved Apr 2008, from http://www.pkv.de.

PMSI. (2000). Programme de Medication des Systemes d'Information. Ministere de l'Emploi de la Solidarite, Paris, France. Retrieved Mar 2008, from http://www.le-pmsi.fr.

Portalfarma. (2001). Base de datos del medicamento y parafarmacia. Consejo General de Colegios Oficiales de Farmaceuticos. Retrieved Apr 2008, from www.portalfarma.com.

Precker, A. S., & Harding, A. (2000). The economics of public and private roles in health care. Insights from Institutional Economics and Organizational Theory. In A. Ron & X. Scheil-Adlung (Eds.), Recent Health Policy Innovations in Social Security, International Social Security Series (Vol. 5, pp. 209-252). London: Transaction Publishers.

Raikou, M., Briggs, A., Gray, A., & McGuire, A. (2000). Centre-specific or average unit costs in multi-centre studies? Some theory and simulation. Health Economics, 9(3), 191-198.

Raisch, D. W., Netravali, S. S., & Harris, C. L. (2003). Pharmacoeconomic implications of large multicenter clinical trials. The Annals of Pharmacotherapy, 37, 447-450.

Ramsay, C. R., Grant, A. M., Wallace, S. A., Garthwaite, P. H., Monk, A. F., & Russell, I. T. (2001). Statistical assessment of the learning curves of health technologies. Health Technology Assessment, 5(12), 1-79.

Rawls, J. (1971). A Theory of Justice. Harvard: Harvard University Press.

Reed, S. D., Anstrom, K. J., Bakhai, A., Briggs, A. H., Califf, R. M., Cohen, D. J., et al. (2005). Conducting economic evaluations alongside multinational clinical trials: Toward a research consensus. American Heart Journal, 149, 434-443.

Register, C., Williams, D., & Bruning, E. (1991). A Note on the efficiency of "switching" hospitals. Quarterly Journal of Business and Economics, 30(3), 3-14.

Reinhardt, U. (1992). Reflections on the meaning of efficiency: can efficiency be separated from equity? Yale Law and Policy Review, 10(2), 302-315.

Retzlaff-Roberts, D., Chang, C. F., & Rubin, R. M. (2004). Technical efficiency in the use of health care resources: a comparison of OECD countries. Health Policy, 69, 55-72.

Revicki, D. A., & Frank, L. (1999). Pharmacoeconomic evaluation in the real world. Effectiveness versus efficacy studies. Pharmacoeconomics, 15(5), 423-434.

RF Ministry of Health. (2002). Branch Standard. The standardization system in the Russian Federation. Health care system. Clinico-economic studies. General provisions.

Rice, N., & Jones, R. (1997). Multilevel models and health economics. Health Economics, 6, 561-575.

Robbins, C. C., Bauer, C. M., Vujicic, N., Gaschler, G. J., Lichty, B. D., Brown, E. G., et al. (2006). Cigarette smoke impacts immune inflammatory responses to influenza in mice. American Journal of Respiratory and Critical Care Medicine, 174(12), 1342-1351.

Roberts, C. (1999). The implication of variation in outcome between health care professionals for the design and analysis of randomised controlled trials. Statistics in Medicine, 18, 2605-2615.

Robinson, J. C. (1997). Physician-hospital integration and economic theory of the firm. Medical Care Research and Review, 54(1), 3-24.

Rosen, R., & Gabbay, J. (1999). Linking health technology assessment to practice. BMJ, 319, 1292-1294.

Rosenbrock, R. (2001). Was ist new public health? Bundesgesundheitsblatt-Gesundheitsforschung–Gesundheitsschutz, 44, 753-762.

Ross, S. A. (1973). The economic theory of agency: The principal's problem. American Economic Review, 63(2), 134-139.

Rote Liste®. (2001). Arzneimittelverzeichnis für Deutschland. Aulendorf, Germany: Rote Liste® Service GmbH Frankfurt/M. Editio Cantor Verlag.

Rugger, J. P., Jarrison, D. T., & Bloom, D. E. (2001). Health and the Economy. In M. H. Merson, R. E. Black & A. J. Mills (Eds.), International Public Health: Diseases, programs, systems, and policies. Gaithersburg, Aspen.

Ruiz, M., Ewig, S., Marcos, M. A., Martinez, J. A., Arancibia, F., Mensa, J., et al. (1999). Etiology of community-acquired pneumonia: impact of age, comorbidity, and severity. American Journal of Respiratory and Critical Care Medicine, 160(2), 397-405.

Rutten-van Molken, M. P., van Doorslaer, E. K., & Till, M. D. (1998). Cost-effectiveness analysis of formoterol versus salmeterol in patients with asthma. Pharmacoeconomics, 14(6), 671-684.

Ryan, M. (1994). Agency in health care: Lessons for economists from sociologists. American Journal of Economics and Sociology, April 53(2), 207-218.

Rychlik, R., Wertheimer, A., Rusche, H., Augustin, M., Nelles, S., & Boecking, W. (2005). Policy decision making and outcomes research in drug utilization. Journal of Public Health, 13, 257-264.

Schreyögg, J. (2004). Gerechtigkeit in Gesundheitssystemen aus ökonomischer Perspektive. Gesundheitswesen, 66, 7-14.

Schulenburg, v. J. M., & Greiner, W. (2000). Health Economics. Tübingen: Mohr-Siebeck.

Schwermann, T., Greiner, W., & Schulenburg, V. D. (2003). Using disease management and market to address the adverse economic effects of drug budgets and price and reimbursement regulations in Germany. Value Health, 6(Suppl 1), 20-30.

Scottish Medicines Consortium (SMC). (2007). Guidance to manufacturers for completion of new product assessment form. Retrieved May 2009, from http://www.scottishmedicines.org.uk/files/.

Sculpher, M. J., Claxton, K., Drummond, M., & McCabe, C. (2006b). Whither trial-based economic evaluation for health care decision making? Health Economics, 15, 677-687.

Sculpher, M. J., & Drummond, M. F. (2006a). Analysis sans frontieres: Can we ever make economic evaluations generalisable across jurisdictions? Pharmacoeconomics, 24(11), 1087-1099.

Sculpher, M. J., Pang, F. S., Manca, A., Drummond, M. F., Golder, S., Urdahl, H., et al. (2004). Generalisability in economic evaluation studies in healthcare: a review and case studies. Health Technology Assessment, 8(49), iii-iv, 1-192.

Sen, A. (1977). Rational fools: a critique of the behavioral foundations of economic theory. Philosophy and Public Affairs, 6, 317-344.

Sen, A. (1985). Commodities and capabilities. Amsterdam: North Holland.

Sen, A. (1987). On Ethics and Economics. Oxford: Basil Blackwell.

Sen, A. (1997). On economic inequality. New York: Oxford University Press.

Shih, Y., & Kauf, T. L. (1999). Reconciling decision models with the real world. An application to anaemia of renal failure. Pharmacoeconomics, 15(5), 481-493.

Skinner, C., Holt, D., & Smith, T. F. (1989). The analysis of complex surveys. New York: Wiley.

Snijders, T. A. B., & Bosker, R. J. (1999). Multilevel analysis: An introduction to basic and advanced multilevel modeling. London: Sage Publishers.

SOIKOS. (2001). Base de datos de costes sanitarios. Barcelona, Spain.

Spath, H.-M., Carrere, M.-O., Fervers, B., & Philip, T. (1999). Analysis of the eligibility of published economic evaluations for transfer to a given health care system: Methodological approach and application to the French health care system. Health Policy, 49(3), 161-177.

Spiegelhalter, D., Best, N., Carlin, B., & Linde, A. (2002). Bayesian measures of model complexity and fit. Journal Of The Royal Statistical Society Series B, 64(4), 583-639.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). WinBUGS User Manual. Version 1.4. [Electronic Version]. Retrieved Dec, 2008 from http://www.mrc-bsu.cam.ac.uk/bugs.

Spiegelhalter, D. J., Thomas, A., Best, N., & Gilks, W. (1996). BUGS 0.5: Bayesian inference using Gibbs sampling [Electronic Version]. Retrieved Jan 2010 from www.mrc-bsu.cam.ac.uk/bugs.

StataCorp. (2005). Statistical software. Release 9.0. College Station, TX: Stata Corporation.

Stigler, G. (1971). The theory of economic regulation. Bell Journal of Economics, The RAND Corporation, 21(1), 3-21.

Stiglitz, J. E. (1974). Incentives and risk sharing in sharecropping. Review of Economic Studies, 41(2), 219-255.

Stiles, R., & So, S. A. (2003). Impact of transaction costs on healthcare outcomes. Expert Rev. Pharmacoeconomics and Outcomes Research, 3(3), 283-291.

Sullivan, S. D., Liljas, B., Buxton, M., Lamm, C. J., O'Byrne, P., Tan, W. C., et al. (2001). Design and analytic considerations in determining the cost-effectiveness of early intervention in asthma from a multinational clinical trial. Controlled Clinical Trials, 22(4), 420-437.

Szende, A., Mogyorósy, Z., Muszbek, N., Nagy, J., Pallos, G., & Dózsa, C. (2002). Methodological guidelines for conducting economic evaluation of healthcare interventions in Hungary: a Hungarian proposal for methodology standards. European Journal of Health Economics, 3, 196-206.

Taghreed, A., Evans, D. B., & Koopmanschap, M. A. (2003). Cost-effectiveness analysis: Can we reduce variability in costing methods. International Journal of Technology Assessment in Health Care, 19(2), 407-420.

Tarmed. (2002). Swiss National Tariff Catalogue. Tarmed Suisse, Bern. Retrieved Mar 2008, from www.tarmed.ch.

Taxordnung. (2002). Taxordnung der Spitäler und der Psychiatrischen Kliniken. Landammann und Regierungsrat des Kantons St.Gallen. Volume 321.61.

The BUGs Project. (2010). DIC: Deviance Information Criterion. Retrieved Feb 2010, from http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/dicpage.shtml.

The Swedish Pharmaceutical Benefits Board. (2008). Working guidelines for the pharmaceutical reimbursement review. Retrieved Mar, 2009, from www.tlv.se/Upload/.../guidelines-pharmaceutical-reimbursement.pdf.

Thomas, D., & Strauss, J. (1998). The micro-foundations of the links between health, nutrition, and development. Journal of Economic Literature, 36, 766-817.

Thompson, K. M. (2002). Variability and uncertainty meet risk management and risk communication. Risk Analysis, 22(8), 647-654.

Thompson, S. G., Nixon, R., & Grieve, R. (2010 forthcoming). Addressing the issues that arise in multicentre cost data, with application to a multinational study. Submitted to Journal of Health Economics.

TIPS. (2001). Tarif Interministeriel des Prestations Sanitaires (TIPS). L'Union des Caisses Nationales de Se´ curite´ Sociale (UCANSS). Paris, France.

Tsuchiya, A., Ikeda, S., Ikegami, N., Nishimura, S., Sakai, I., & Fukuda, T. (2002). Estimating an EQ-5D population value set: the case of Japan. Health Economics, 11, 341-353.

Tuckman, H., & Chang, C. (1988). Cost convergence between for-profit and not-for-profit nursing homes: Does competition matter? Quarterly Review of Economics and Business, 28(4), 50-65.

Tunis, S. L., Johnstone, B. M., Kinon, B. J., Barber, B. L., & Browne, R. A. (2000). Designing naturalistic prospective studies of economic and effectiveness outcomes associated with novel antipsychotic therapies. Value in Health, 3(3), 232-242.

United Nations. (2008). South Africa: Country profile. Retrieved Jul, 2008, from http://www.nationsencyclopedia.com/index.html

Urdahl, H., Manca, A., & Sculpher, M. J. (2006). Assessing generalisability in model-based economic evaluation studies: a structured review in osteoporosis. Pharmacoeconomics, 24(12), 1181-1197.

Vademecum. (2001). Vedemecum Interacional. Especialidades Farmaceuticas y Biologicas, Productos Y Articulos De Parafarmacia, Metodos De Diagnostico. In. Madrid, Spain.

van den Berg, S. M., Beem, L., & Boomsma, D. I. (2006). Fitting genetic models using Markov Chain Monte Carlo algorithms with BUGS. Twin Research and Human Genetics, 9(3), 334-342.

van Walbeek, C. (2002). Recent trends in smoking prevalence in South Africa. South African Medical Journal, 92(6), 468-472.

Vidal-Semp. (2001). Thesaurus Vidal Semp - CD-Rom. Paris: Vidal.

Wagstaff, A., Doorslaer, E. V., Hattem, V. D. B., Calonge, S., Christiansen, T., Citoni, G., et al. (1999). Equity in the finance of health care: Some further international comparisons. Journal of Health Economics, 18, 263-290.

Walter, E., & Zehetmayr, S. (2006). Guidelines on health economic evaluation. Institute for pharmaeconomic research. Consensus Paper. Retrieved Mar, 2009, from http://www.ispor.org/peguidelines/source/Guidelines_Austria.pdf.

Welte, R., Feenstra, T., Jager, H., & Leidl, R. (2004). A decision chart for assessing and improving the transferability of economic evaluation results between countries. Pharmacoeconomics, 22(13), 857-876.

WHO. (1952). Expert committee on public-health administration. World Health Organization. Tech Rep Ser No. 51(55)), 1-41.

WHO. (1986). Ottawa Charter for Health Promotion. World Health Organization [Electronic Version]. Retrieved Nov 2009 from http://www.who.int/hpr/NPH/docs/ottawa_charter_hp.pdf.

WHO. (2009). European Health for All database (HFA-DB). Retrieved Sep, 2009, from http://www.euro.who.int/hfadb

WHO European Centre for Health Policy. (2008). European observatory on health systems and policies. Retrieved Jan 2009, from http://www.euro.who.int/observatory.

Wilke, R. J. (2003). Tailor-made or off-the-rack? The problem of transferability of health economic data. Expert Review in Pharmacoeconomics and Outcomes Research, 3(1), 1-4.

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

Willan, A. R., & Kowgier, M. (2007). Cost-effectiveness analysis of a multinational RCT with a binary measure of effectiveness and an interacting covariate. Health Economics, 17(7), 777-791.

Willan, A. R., Pinto, E. M., O'Brien, J. B., Kaul, P., Goeree, R., Lynd, L., et al. (2005). Country specific cost comparisons from multinational clinical trials using

empirical Bayesian shrinkage estimation: the Canadian ASSENT-3 economic analysis. Health Economics, 14(4), 327-338.

Williamson, O. E. (1973). Markets and hierarchies: Some elementary considerations. American Economic Association, 63(2), 316-334.

Williamson, O. E. (1975). Markets and hierarchies: Analysis and antitrust implications. New York: The Free Press.

Williamson, O. E. (1985). The Economic Institutions of Capitalism. New York: The Free Press.

Witter, S., Ensor, T., Jowett, M., & Thompson, R. (2000). Health economics for developing countries: a practical guide. Basingstoke: Macmillan.

Woodhead, M., Fleming, D., & Wise, R. (2004). Antibiotics, resistance, and clinical outcomes. BMJ, 328(7451), 1270-1271.

Wu, Y. W. B. (1995). Hierarchical linear models: a multilevel data analysis technique. Nursing Research, 44, 123.

Appendix A. Assessment of Multicollinearity

5 countries sample:

```
pwcorr totallos cost x4age smoknum y3smok x2cap, sig sidak
```

	totallos	cost	x4age	smoknum	y3smok	x2cap
totallos	1.0000					
cost	0.6610	1.0000				
	0.0000					
x4age	0.3176	0.2352	1.0000			
	0.0000	0.0002				
smoknum	0.1227	0.0872	-0.2158	1.0000		
	0.3170	0.8328	0.0011			
y3smok	-0.0039	0.0338	-0.2505	0.8031	1.0000	
	1.0000	1.0000	0.0001	0.0000		
x2cap	0.1104	0.1870	0.1160	0.0428	-0.0066	1.0000
	0.4918	0.0090	0.4081	0.9998	1.0000	

10 countries sample:

```
pwcorr totallos x4age smoknum x2cap y3smok, sig sidak
```

	totallos	x4age	smoknum	x2cap	y3smok
totallos	1.0000				
x4age	0.3376	1.0000			
	0.0000				
smoknum	0.0762	-0.2331	1.0000		
	0.4465	0.0000			
x2cap	0.0845	0.0242	-0.0249	1.0000	
	0.2999	0.9996	0.9995		
y3smok	-0.0367	-0.2855	0.8052	-0.0386	1.0000
	0.9887	0.0000	0.0000	0.9834	

Bivariate correlation of dependent variable LOS and covariates age, smoking and severity of CAP at admission does not indicate strong correlation and therefore potential multicollinerity between covariates. Thus, all covariates can be used in hierarchical models.

Appendix B. WinBUGS code for HM 3

```
model
{
for(i in 1:Npatients)
{
effectiveness[i] ~ dbern(meanE[i])
alpha[i] <- muES+gammaE[country[i]]
tau[i] <- (capital.deltaE+deltaE[country[i]])*j[i] +(thetaEX4+omegaEX4*j[i])*(X4AGE[i]-
X4AGE.bar)
meanE[i] <- min(max(alpha[i], -tau[i]), 1 - tau[i]) + tau[i]
cost[i] ~ dgamma(shapeC[country[i]], rateC[i])
rateC[i] <- shapeC[country[i]]/meanC[i]
meanC[i] <- muCS*(1-j[i])+muCT*j[i]+gammaC[country[i]]+deltaC[country[i]]*j[i]
+(thetaCX2+omegaCX2*j[i])*X2CAP[i]+(thetaCX4+omegaCX4*j[i])*(X4AGE[i]-
X4AGE.bar)+(thetaCY3+omegaCY3*j[i])*(Y3SMOK[i]-
Y3SMOK.bar)+beta[j[i]+1]*(effectiveness[i] - meanE[i])
dC[i] <- -loggam(shapeC[country[i]])+shapeC[country[i]]*log(rateC[i])
+(shapeC[country[i]]-1)*log(cost[i])-rateC[i]*cost[i]
dE[i] <- log(pow(meanE[i],effectiveness[i])
*pow(1-meanE[i],1-effectiveness[i]))
}
for(k in 1:M)
{
deltaEE[k] ~ dnorm(0, tau.deltaE)
gammaEE[k] ~ dnorm(0, tau.gammaE)
gammaCC[k] ~ dnorm(0, tau.gammaC)
deltaCC[k] ~ dnorm(0, tau.deltaC)
deltaE[k] <- deltaEE[k] - (inprod(country.n[], deltaEE[])/sum(country.n[]))
gammaE[k] <- gammaEE[k] - (inprod(country.n[], gammaEE[])/sum(country.n[]))
gammaC[k] <- gammaCC[k] - (inprod(country.n[], gammaCC[])/sum(country.n[]))
deltaC[k] <- deltaCC[k] - (inprod(country.n[], deltaCC[])/sum(country.n[]))
}
# Overall parameters
capital.deltaCcov <- muCT - muCS
capital.deltaEcov <- capital.deltaE
EC <- capital.deltaEcov*capital.deltaCcov # used for estimating C(De, Dc) for X4AGE = 0
for(i in 1:101) {ceac0[i] <- step(capital.deltaEcov*1000*(i - 1) - capital.deltaCcov)}
ProbSE <- step(capital.deltaEcov)*step(-capital.deltaCcov)
# Country 1's parameters
capital.deltaE.c1 <- capital.deltaEcov+deltaE[1]
capital.deltaC.c1 <- capital.deltaCcov+deltaC[1]
EC.c1 <- capital.deltaE.c1*capital.deltaC.c1 # used for estimating C(De, Dc) for X4AGE = 0
for(i in 1:101) {ceac.c1[i] <- step(capital.deltaE.c1*1000*(i - 1) - capital.deltaC.c1)}
ProbSE.c1 <- step(capital.deltaE.c1)*step(-capital.deltaC.c1)
# Country 2nd parameters
capital.deltaE.c2 <- capital.deltaEcov+deltaE[2]
capital.deltaC.c2 <- capital.deltaCcov+deltaC[2]
EC.c2 <- capital.deltaE.c2*capital.deltaC.c2 # used for estimating C(De, Dc) for X4AGE = 0
for(i in 1:101) {ceac.c2[i] <- step(capital.deltaE.c2*1000*(i - 1) - capital.deltaC.c2)}
ProbSE.c2 <- step(capital.deltaE.c2)*step(-capital.deltaC.c2)
# Country 3rd parameters
capital.deltaE.c3 <- capital.deltaEcov+deltaE[3]
```

```
capital.deltaC.c3 <- capital.deltaCcov+deltaC[3]
EC.c3 <- capital.deltaE.c3*capital.deltaC.c3 # used for estimating C(De, Dc) for X4AGE = 0
for(i in 1:101) {ceac.c3[i] <- step(capital.deltaE.c3*1000*(i - 1) - capital.deltaC.c3)}
ProbSE.c3 <- step(capital.deltaE.c3)*step(-capital.deltaC.c3)
# Country 4th parameters
capital.deltaE.c4 <- capital.deltaEcov+deltaE[4]
capital.deltaC.c4 <- capital.deltaCcov+deltaC[4]
EC.c4 <- capital.deltaE.c4*capital.deltaC.c4 # used for estimating C(De, Dc) for X4AGE = 0
for(i in 1:101) {ceac.c4[i] <- step(capital.deltaE.c4*1000*(i - 1) - capital.deltaC.c4)}
ProbSE.c4 <- step(capital.deltaE.c4)*step(-capital.deltaC.c4)
# Country 5th parameters
capital.deltaE.c5 <- capital.deltaEcov+deltaE[5]
capital.deltaC.c5 <- capital.deltaCcov+deltaC[5]
EC.c5 <- capital.deltaE.c5*capital.deltaC.c5 # used for estimating C(De, Dc) for X4AGE = 0
for(i in 1:101) {ceac.c5[i] <- step(capital.deltaE.c5*1000*(i - 1) - capital.deltaC.c5)}
ProbSE.c5 <- step(capital.deltaE.c5)*step(-capital.deltaC.c5)
X4AGE.bar<-mean(X4AGE[])
Y3SMOK.bar<-mean(Y3SMOK[])
muES ~ dunif(0, 1)
capital.deltaE ~ dunif(-1, 1)
for(k in 1:M) {shapeC[k] ~ dunif(0, 30)}
muCS ~ dunif(300, 25000)
muCT ~ dunif(300, 25000)
omegaCX2~ dnorm(0, 1.0E-2)
omegaCX4~ dnorm(0, 1.0E-2)
omegaCY3~ dnorm(0, 1.0E-2)
omegaEX4~ dnorm(0, 1.0E-2)
thetaCX2~ dnorm(0, 1.0E-2)
thetaCX4~ dnorm(0, 1.0E-2)
thetaCY3~ dnorm(0, 1.0E-2)
thetaEX4~ dnorm(0, 1.0E-2)
for(i in 1:2) {beta[i] ~ dnorm(0, 1.0E-2)}
tau.deltaE <- 1/ss.deltaE
ss.deltaE <- exp(lnss.deltaE)
s.deltaE <- sqrt(ss.deltaE)
lnss.deltaE ~ dunif(-10, 5)
tau.gammaE <- 1/ss.gammaE
ss.gammaE <- exp(lnss.gammaE)
s.gammaE <- sqrt(ss.gammaE)
lnss.gammaE ~ dunif(-10, 5)
tau.gammaC <- 1/ss.gammaC
ss.gammaC <- s.gammaC*s.gammaC
s.gammaC ~ dunif(1,4000)
tau.deltaC <- 1/ss.deltaC
ss.deltaC <- s.deltaC*s.deltaC
s.deltaC~dunif(100,4000)
dev <- -2*(sum(dC[])+sum(dE[]))
}
```

Appendix C. Histogram of LOS and Costs

Figure 18: Cost Histogram for Treatment and Comparator Arms Combined

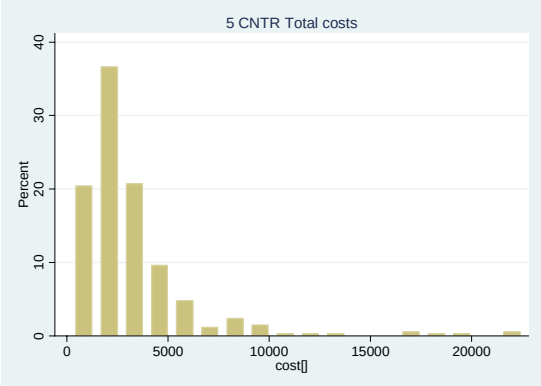


Figure 19: Cost Histogram for Treatment Arm

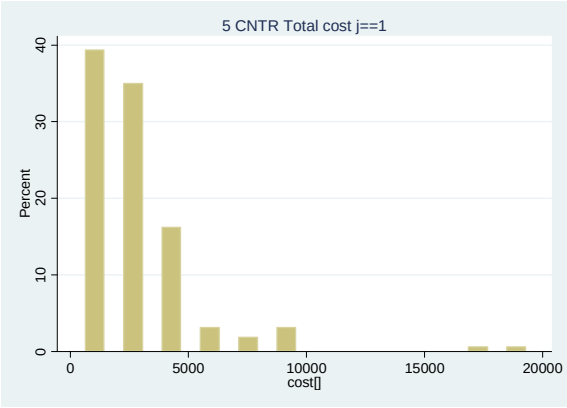


Figure 20: Cost Histogram for Comparator Arm

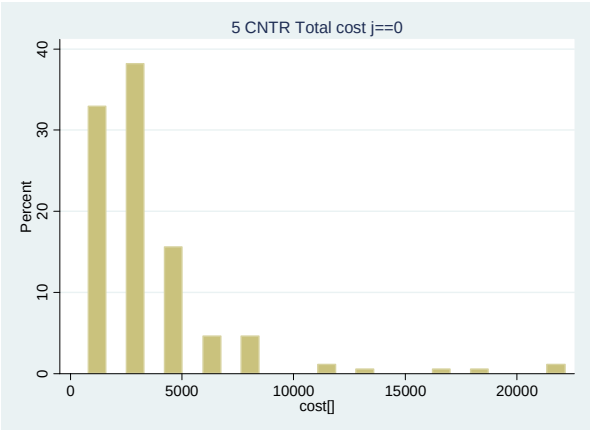


Figure 21: Cost Histogram for Treatment and Comparator Arms per Country

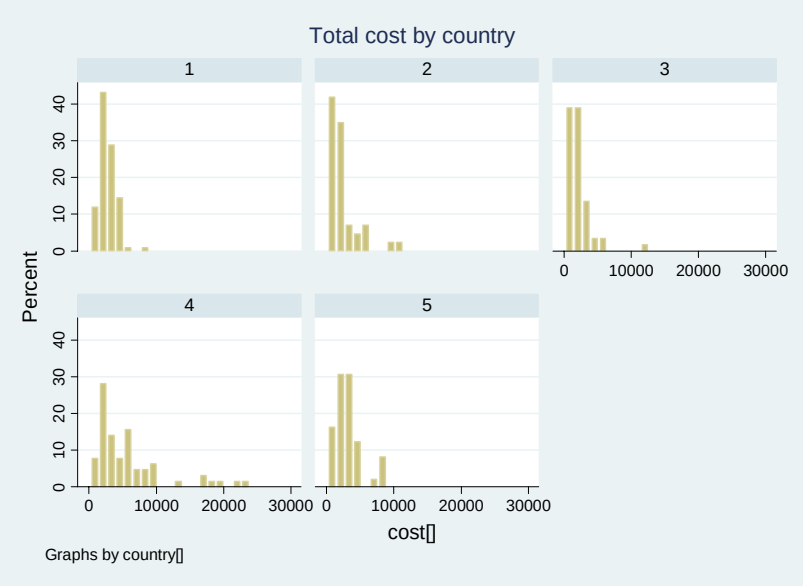


Figure 22: LOS Histogram for Treatment and Comparator Arms Combined

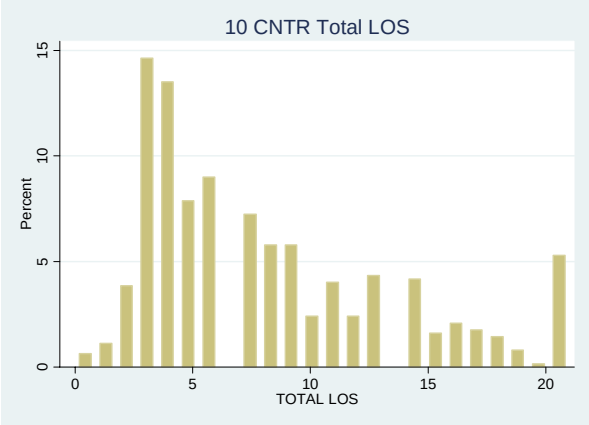
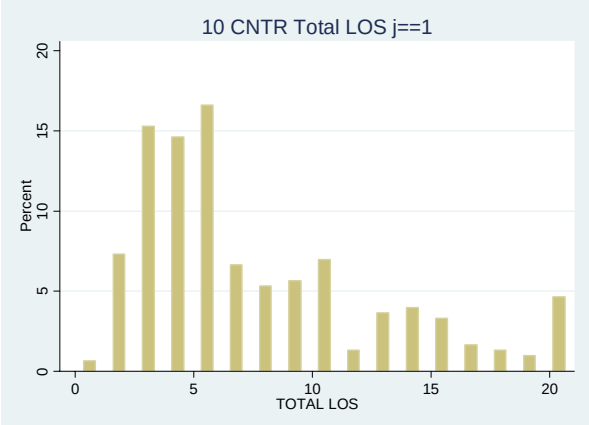


Figure 23: LOS Histogram for Treatment Arm



Appendix D. Costing Methodology Applied In Case Study

Table 33: Case Study Resource Costing: Study Medication

	Route	Cost Unit	In Hospital Cost/EUR	Out Hospital Cost/EUR	Source	Comment
France						
Moxifloxacin	IV	400 mg	46.85	30.45	BSP information	Price at market introduction
Moxifloxacin	PO	400 mg	5.18	3.37	BSP information	Official market price
Co-amoxiclav	IV	1.2 g	4.21	2.73		Official public selling price of Augmentin
Co-amoxiclav	PO	625 mg	0.91	0.59		Official public selling price of Augmentin
Clarithromycin	IV	500 mg	17.53	11.39		Official public selling price of Klacid iv
Clarithromycin	PO	500 mg	2.00	1.30	Vidal-Semp, 2001	Official public selling price of Klacid/Clarithomycin PO
Germany						
Moxifloxacin	IV	400 mg	60.83	59.91	BSP information	Price at market introduction
Moxifloxacin	PO	400 mg	5.65	5.19	BSP information	Official market price
Co-amoxiclav	IV	1.2 g	11.28	10.36		Official selling price of Augmentan out of pharmacy
Co-amoxiclav	PO	625 mg	3.00	2.77		Weighted average based on market share: 33% Augmentan and 66% generics (IMS Health, 2002)
Clarithromycin	IV	500 mg	31.59	30.67		IV form not available in Germany; cost assumed to be 8.8 times PO cost as per corresponding ratio in France
Clarithromycin	PO	500 mg	3.59	3.36	Rote Liste®, 2001	Weighted average based on market share: 60% Klacid, 5% Biaxin, 35% reimported Klacid (IMS Health, 2002)
Spain						
Moxifloxacin	IV	400 mg	40.93	31.11	BSP information	Price at market introduction
Moxifloxacin	PO	400 mg	3.61	2.75	BSP information	Official market price
Co-amoxiclav	IV	1.2 g	3.18	2.42		Based on price of Augmentine 100% market share (IMS Health, 2002)
Co-amoxiclav	PO	625 mg	0.40	0.30		Weighted average based on market share: 67% Augmentine, 15% Duonasa, 12% Eupeclanic, 3% Clavumox, 3% Clavucid (IMS Health, 2002)
Clarithromycin	IV	500 mg	13.71	10.42	Vademecum, 2001;	Based on price of Klacid and Bremon
Clarithromycin	PO	500 mg	1.69	1.29	Portalfarma, 2001	Based on price of Klacid and Bremon

	Route	Cost Unit	In Hospital Cost/EUR	Out Hospital Cost/EUR	Source	Comment
Switzerland						
Moxifloxacin	IV	400 mg	60.61	54.55	BSP information	Price at market introduction
Moxifloxacin	PO	400 mg	4.74	4.27	BSP information	Official market price
Co-amoxiclav	IV	1.2 g	19.05	17.15		Official public selling price of Augmentin
Co-amoxiclav	PO	625 mg	1.87	1.68		Official public selling price of Augmentin
Clarithromycin	IV	500 mg	29.04	26.14		Official public selling price of Klacid iv
Clarithromycin	PO	500 mg	1.93	1.74	Documed, 2001	Official public selling price of Klacid/Clarithomycin PO
UK						
Moxifloxacin	IV	400 mg	51.41	51.41	BSP information	Price at market introduction
Moxifloxacin	PO	400 mg	3.43	3.43	BSP information	Official market price
Co-amoxiclav	IV	1.2 g	4.66	4.66		Based on price of Augmentin 100% market share (IMS Health, 2002)
Co-amoxiclav	PO	625 mg	1.17	1.17	British National Formulary, 2001; NHS Drug Tariff, 2001	Weighted average based on market share: 71% Augmentin, 29% Non-branded (IMS Health, 2002)
Clarithromycin	IV	500 mg	17.55	17.55		Based on price of Klaricid 100% market share (IMS Health, 2002)
Clarithromycin	PO	500 mg	2.77	2.77		Based on price of Klaricid 100% market share (IMS Health, 2002)

Table 34: Case Study Resource Costing: Concomitant Medication

	Data Source	Comment
The complete list with unit prices for concomitant medications available on request		
France	Vidal-Semp, 2001	Reimbursement rate of 65% was applied
Germany	Rote Liste®, 2001	Patient co-payment of 4.60 EUR (for medium sized pack) was assumed
Spain	Vademecum, 2001; Portalfarma, 2001	Reimbursement rate of 76% was applied
Switzerland	Documed, 2001	Reimbursement rate of 90% was applied
UK	British National Formulary 2001; NHS Drug Tariff, 2001	Reimbursement rate of 100% was applied, medication prices include a wholesaler margin of 12.5%

Table 35: Case Study Resource Costing: Hospitalization and Radiological Procedures

Ward Type	Unit	Unit/EUR	Source	Comment
France				
General	1 day	419.45		
Intensive Care	1 day	1058.07		Cost for 1999 updated to 2000 by factor 1.02; 4% has been removed representing cost of study
All Others	1 day	419.45	PMSI, 2000	medications; costs of other ward types assumed to be equal to general ward
Germany				
General	1 day	201.10		Cost for 1999 updated to 2000 using the consumer price index for health care (113.6/107.6); 4% has been
Intensive Care	1 day	775.21		removed representing cost of study medications, costs of other ward types assumed to be equal to general
All Others	1 day	201.10	PKV, 1999/2000	ward
Spain				
General	1 day	286.08		
Intensive Care	1 day	1055.04		4% has been removed representing cost of study medications, costs of other ward types assumed to be
All Others	1 day	286.08	SOIKOS, 2001	equal to general ward
Switzerland				
General	1 day	256.89		
Intensive Care	1 day	1262.01		Cost for 1999 updated to 2000 using the consumer price index for health care (113.6/107.6); 4% has been
All Others	1 day	256.89	Taxordnung, 2002	removed representing cost of study medications

Ward Type	Unit	Unit/EUR	Source	Comment
UK				
General	1 day	237.56	NHS Reference costs, 2001/2002	Cost represents mean per diem for non-elective patients with bronchopneumonia; 4% has been substracted
Intensive Care	1 day	1795.11		Cost represents mean cost for all types of admissions; 4% has been substracted
All Others	1 day	237.56		Assumed General ward cost for all other ward types

	Cost Unit	Unit/EUR	Source	Comment
France				
Chest X-Ray	1 X-ray	17.24		
CT Scan - Chest	1 CT scan	94.08	NGAP, 2001	A reimbursement rate of 70% was applied to radiological procedures.
Germany				
Chest X-Ray	1 X-ray	16.32		
CT Scan - Chest	1 CT scan	134.06	GOA, 2000	No co-payments were applied to these costs
Spain				
Chest X-Ray	1 X-ray	18.00		
CT Scan - Chest	1 CT scan	130.00	SOIKOS, 2001	
Switzerland				
Chest X-Ray	1 X-ray	58.41	Tarmed, 2002	Reimbursement rate of 90% was applied
CT Scan - Chest	1 CT scan		NA	
UK				
Chest X-Ray	1 X-ray	20.38		
CT Scan - Chest	1 CT scan	73.67	NHS Drug Tariff, 2001	Reimbursement rate of 100% was applied

Table 36: Case Study Resource Costing: Therapeutic Adjuncts

Procedure	Cost Unit	Healthcare System Cost/Unit	Source	Comment
France				Reimbursement rate of 65% was applied
Aerosol Breathing Therapy	1 session	0.86		Without a humidifier, not including medication
Chest Physiotherapy	20 minutes	1.84		Assumed 5 visits per week
Inhalation Therapy	1 session	5.97		Assumed therapy for <12 hours/day
Nasal C-Pap	1 session	6.34		Cost includes C-PAP and oxygen
Oxygen (Face Mask)	1 session	4.81	NGAP, 2001	Assumed short-term treatment (< 4 weeks)
Oxygen (Nasal Cannula)	1 session	4.51	TIPS, 2001	Assumed short-term treatment (< 4 weeks)
Germany				Reimbirsement is item-specific
Aerosol Breathing Therapy	1 session	9.87		Cost for general physical therapy from West Germany; co-pay of 15% has been removed
Chest Physiotherapy	20 minutes	9.87	Physiotherapy offices in Northrhein-Westphalia, 2001	Cost for general physical therapy from West Germany; co-pay of 15% has been removed
Inhalation Therapy	1 session	13.08		Cost includes therapy and inhalation; cost from West Germany; 15% co-pay removed
Nasal C-Pap	1 session	0.00		Not reimbursed
Oxygen (Face Mask)	1 session	0.00		Not reimbursed
Oxygen (Nasal Cannula)	1 session	0.00		Not reimbursed
Spain				Reimbursement rate of 100% was applied
Aerosol Breathing Therapy	1 session	3.00		Assumed 5 sessions per week
Chest Physiotherapy	20 minutes	0.00		Not reimbursed
Inhalation Therapy	1 session	4.00		Cost not found; assumed same cost as oxygen (face mask)
Nasal C-Pap	1 session	4.00		Cost not found; assumed same cost as oxygen (face mask)
Oxygen (Face Mask)	1 session	4.00		Assumed 5 session per week
Oxygen (Nasal Cannula)	1 session	4.00	SOIKOS, 2001	Cost not found; assumed same cost as oxygen (face mask)
Switzerland				Reimbursement rate of 90% was applied
Aerosol Breathing Therapy	1 session	29.96	Physioswiss, 2001	Allgemeine Physiotherapie, Cost for 1999 updated to 2000 using the consumer price index for health care (113.6/107.6)
Chest Physiotherapy	20 minutes	29.96		
Inhalation Therapy	1 session	29.96		
Nasal C-Pap	1 session	29.96		

Procedure	Cost Unit	Healthcare System Cost/Unit	Source	Comment
Oxygen (Face Mask)	1 session	29.96		
Oxygen (Nasal Cannula)	1 session	29.96		
UK				Reimbursement rate of 100% was applied
Aerosol Breathing Therapy	1 session	25.08		Clinic visit to a physiotherapist; assumed 5 sessions per week
Chest Physiotherapy	20 minutes	25.08		Clinic visit to a physiotherapist; assumed 5 sessions per week
Inhalation Therapy	1 session	25.08		Clinic visit to a physiotherapist; assumed 5 sessions per week
Nasal C-Pap	1 session	11.30		Cost not found; assumed same cost as oxygen face mask
Oxygen (Face Mask)	1 session	11.30	NHS Drug Tariff, 2001	Set up and supplies provided through pharmacist; assumed cost of 1 set paid for over 1 year plus 1 cyclinder of oxygen per day
Oxygen (Nasal Cannula)	1 session	11.30		Cost not found; assumed same cost as oxygen face mask

Note: unit costs data was available for UK, France, Germany and Spain (BSP study report, 2000), unit costs data for Switzerland was collected by the author.

Declaration

I hereby certify that this thesis is result of my own work, investigation and analysis, if not otherwise stated. External sources are acknowledged by citation and included into bibliography.

I herewith confirm, that this thesis was not previously submitted elsewhere. Furthermore, I have not attempted and was not admitted to doctoral studies in other academic institutions.

Anna Filonenko Berlin, 01.06.2010