A Jean-Charles,

A Clémentine,

Remerciements

A toutes les personnes qui ont, de près ou de loin, contribué à l'aboutissement de ce travail.

Mes remerciements s'adressent tout d'abord à mes promoteurs, le professeur Yves Horsmans, le docteur Claire Beguin et le professeur Marie-Christine Closon. Je leur suis reconnaissante pour la confiance qu'ils ont placée en moi et le soutien constant qu'ils m'ont donné pour mener à bien cette recherche. Merci pour leur porte toujours ouverte, leurs suggestions et nombreuses relectures, leur discernement et l'autonomie accordée.

Je voulais également remercier le docteur Ralph Crott pour ses nombreux apports méthodologiques tout au long de ma thèse. Merci pour son enthousiasme et sa disponibilité.

Je remercie encore les autres membres de mon comité d'encadrement et du jury pour avoir accepté cette invitation. Merci aux Professeurs William d'Hoore, Robert Launois et Hugo Robays pour leurs conseils judicieux, leurs commentaires constructifs et le temps qu'ils m'ont consacré.

Merci aussi aux membres des cliniques universitaires Saint-Luc, aux membres du centre fédéral d'expertise des soins de santé (KCE) et autres membres externes qui ont collaboré à ce projet, et tout particulièrement à Irina Cleemput et Lieven Annemans pour leur apport méthodologique. Je remercie aussi à la firme Roche pour ses relectures et conseils lors de la deuxième partie de ma thèse.

Merci également à tous mes collègues du SESA-CIES et autres membres de l'ESP pour ces quatre années passées en leur compagnie. Merci à Maria-Isabel, Laurence, Evelien et Julie pour leurs nombreux conseils, notamment en matière statistique. Merci à Isabelle R, Isabelle P et à Gabriel

B pour leur aide quant à l'utilisation des bases de données. Merci au docteur Etienne De Clercq de m'avoir fait profiter de sa récente expérience en tant que thésard. Merci à Jean-Bernard et à Mostafa pour le support informatique si souvent nécessaire. Merci à Nathalie, Catherine et Muriel pour leur aide administrative. Merci encore à Yolande, Maria-Isabel, Isabelle R, Isabelle V, Isabelle P, Caroline, Micheline, Charlotte, Geoffroy, Fati, Yseult et tous les autres pour leur soutien constant et pour m'avoir apporté de la bonne humeur quand le moral baissait.

Enfin, merci à toutes les personnes qui me sont chères et sans qui, il aurait été difficile d'avancer. Merci du fond du cœur à mon mari, Jean-Charles Byl, pour son soutien sans faille, sa présence constante et toute l'attention qu'il me donne. Merci pour ses encouragements permanents face à mes doutes, pour sa patience et pour sa présence à mes côtés dans les bons comme dans les mauvais moments. Sans lui, je ne serais jamais arrivée aussi loin. Merci aussi à mes parents et à ma sœur pour leur soutien, leur aide et pour tout ce qu'ils m'ont apporté depuis toujours. Merci encore à mes beaux-parents pour leurs multiples encouragements.

Enfin, merci à ma petite Clémentine pour la force qu'elle m'a apportée à la fin de ma thèse. L'aboutissement de cette thèse restera pour toujours indissociable à sa naissance.

TABLE OF CONTENTS

INTRODUCTION **7**

1 BACKGROUND AND DEFINITION **7**

2 METHODS AND CONCERNS IN THE FIELD OF ECONOMIC EVALUATIONS **10**

2.1 PRIMARY ECONOMIC EVALUATIONS 10

2.1.1 The limitation of randomized clinical trials 10

2.1.2 The contribution of observational and modelling studies 14

2.2 REVIEWS OF EXISTING ECONOMIC EVALUATIONS 18

2.2.1 Quality assessment instruments in the field of clinical studies and meta-analyses 18

2.2.2 Quality assessment of economic evaluations 20

3 OBJECTIVES AND OUTLINE OF THE THESIS **21**

PART 1 : THE USE OF ADMINISTRATIVE DATABASES **31**

1 METHODOLOGICAL CHOICES AND RESEARCH QUESTIONS . **31**

2 ASSESSING THE QUALITY OF PHARMACOLOGICAL TREATMENTS FROM ADMINISTRATIVE DATABASES: THE CASE OF LOW-MOLECULAR-WEIGHT HEPARIN AFTER MAJOR ORTHOPEDIC SURGERY **37**

3 CONCLUSION AND PERSPECTIVES **49**

PART 2 : THE USE OF MARKOV MODEL 55

1 METHODOLOGICAL CHOICES AND RESEARCH QUESTION.... 55

2 A HEALTH ECONOMIC MODEL TO ASSESS THE COST-EFFECTIVENESS OF PEG IFN α-2A AND RIBAVIRIN IN PATIENTS WITH MILD CHRONIC HEPATITIS C 59

3 A HEALTH ECONOMIC MODEL TO ASSESS THE COST-EFFECTIVENESS OF PEGYLATED INTERFERON α-2A AND RIBAVIRIN IN PATIENTS WITH MODERATE CHRONIC HEPATITIS C AND PERSISTENTLY NORMAL ALANINE AMINOTRANSFERASE LEVELS..... 75

4 CONCLUSION AND PERSPECTIVES 89

PART 3 : THE USE OF QUALITY ASSESSMENT INSTRUMENTS 97

1 METHODOLOGICAL CHOICES AND RESEARCH QUESTIONS . 97

2 COMPARISON OF THREE INSTRUMENTS ASSESSING THE QUALITY OF ECONOMIC EVALUATIONS: A PRACTICAL EXERCISE ON ECONOMIC EVALUATIONS OF THE SURGICAL TREATMENT OF OBESITY..... 101

3 CONCLUSION AND PERSPECTIVES 115

DISCUSSION..... 121

NOTES..... 129

INTRODUCTION

1 BACKGROUND AND DEFINITION

Health care expenditures have been growing much faster than the gross domestic product in the last years and are expected to continue to rise (1;2). In this context, curbing such a growth has become essential for policy makers. Nevertheless, the control of health care expenditures can not be achieved to the detriment of health care access and health care quality. The priority would be to reduce health care costs without deteriorating the quality of health care, or to improve the quality of health care at a reasonable cost (3). Consequently, the routinely use of full economic evaluations, i.e. analyses comparing both costs and outcomes of at least two health care programs (definition Drummond *et al* (3)) should be promoted.

Three types of full economic evaluations exist: cost-effectiveness analyses, cost-utility analyses and cost-benefit analyses. These economic evaluations differ in the way they measure outcomes of compared programs. In cost-effectiveness analyses, outcome data are quantitative and are defined in terms of final endpoint such as life years gained; or of intermediary outcomes such as the decrease in blood pressure. In cost-utility analyses, outcome data are measured in terms of life years gained adjusted by the quality of life, using a measure such as quality-adjusted life years (QALYs). Thus, this kind of analysis takes both morbidity and mortality into account. Finally, in cost-benefit analysis, outcomes are transformed in monetary value, representing the people willingness to pay (2;3).

The major aim of economic evaluations is to inform health care decision makers in order to help them to allocate scarce resources in a most efficient way. In other words, the budget allocated to health care is limited and the use of economic evaluations could help policy makers in the definition of priorities established in order to maximize people's health and quality of life, and this, by making the best use of available resources.

In practice, decision makers should thus use economic evaluations to prioritize health care programs for which outcomes are as good as or better than outcomes with other health care programs for the same amount of money invested. Moreover, prioritizations should not only be based on comparisons of health care programs within a specific disease but also between diseases. Indeed, if the Belgian Drug Reimbursement Committee decides to reimburse the most efficient treatment for a specific disease, such a choice will have an impact on the decision to reimburse or not treatments of other diseases, the budget being limited. So, the setting of prioritization should be based on comparisons of health care programs for every disease area and both economic evaluations and budget impact analyses should be taken into account in the decision process.

However, many discussions on the insufficient use of economic evaluations in the health care decision process can be found in the literature. Among the major causes, the lack of generalizability to other settings and the lack of real world data are often cited. Economic evaluations are mostly conducted abroad and more especially in United States (US), limiting the transferability of results to other settings (4). Consequently, when performing a systematic review in order to make decision on the reimbursement of a health care program in Belgium, a specific discussion on how foreign costs

and outcomes data will vary according to the Belgian setting is important.

Moreover, health care decision makers have often not enough expertise too assess if the study was of good quality and thus prefer not to use them (4). They also have major concerns about the large number of assumptions and the lack of transparency in modelling studies and consider them as a “black box” (4;5). Then, the few number and poor quality of the majority of economic evaluations was criticized, especially concerning the methodologies used to assess cost data. Indeed, methods to measure and valued costs are often unclear and do not include all relevant costs in the societal perspective, as advised in the guidelines (6;7). The need for assessing the quality of economic evaluations is thus highlighted by health care decision makers and some of them were interested in the implementation of a global quality score to only take into account studies of high quality (4).

The large variety of methods used in economic evaluations was also criticized and has lead to the development of guidelines to standardize the methods (8-10). However, methodological improvements are yet needed. To determine how to improve the information available to Belgian health care decision makers, we thus analyzed which methods could be used in the field of economic evaluations and focused on their major concerns. Moreover, because the improvement of the quality of clinical studies has been examined for a longer time, we also investigated methodological concerns in the field of economic evaluations at the light of concerns and standards in the clinical field.

2 METHODS AND CONCERNS IN THE FIELD OF ECONOMIC EVALUATIONS

In economic evaluations, two levels of evidences are available: primary economic evaluations and reviews of existing economic evaluations, pooling primary economic evaluations. Improving the information available for health care decision makers can be done at these two levels.

2.1 Primary economic evaluations

Data to estimate the benefits and costs of an intervention can be retrieved from various sources such as randomized clinical trial, non-randomized clinical trials, observational studies, interviews or expert opinions (11;12). In the field of clinical studies, randomized controlled trials (RCTs) are considered as the gold standard, allowing to minimize bias and to control for confounding factors (12;13). However, because RCTs have some limitations, they can not be the only source of data in economic evaluations to completely inform health care decision makers.

2.1.1 The limitation of randomized clinical trials

The major weaknesses of RCTs when used in economic evaluations concern the lack of external validity, the lack of relevant comparators, the short time horizon, the insufficient sample size, and issues concerning ethical and cost aspects (14;15).

2.1.1.1 Lack of external validity and lack of effectiveness data

Internal validity of a study can be defined by the degree to which the cause and effect relationship between the treatment and the observed outcome are likely to be true, so that the observed differences between groups is due to the effect of the intervention and not to other causes such as the protocol of the study. On the other hand, the external validity of a study concerns the degree to which results can be generalizable to other patient populations or to other settings (16). The design of RCTs implies a high internal validity but a low external validity. Indeed, processes such as highly precise and controlled protocol, blinding, and strict patients' eligibility criteria allows to reduce bias and to control for confounding factors but limit the generalizability of results.

In RCTs, outcomes are measured in a controlled environment and their protocol does not always follow the daily clinical practice. They usually have more frequent physician visits and more laboratory tests than in daily clinical practice, which is one of the most important challenges for economic evaluations, known by the name of "protocol-driven care" costs and effects. For example, in RCTs assessing prophylaxis against venous thromboembolism (VTE), events are systematically detected by venography, which is not a systematic procedure in practice. So, firstly, the costs of examination are higher than in clinical practice and secondly, more events are identified. Indeed, symptomatic but also asymptomatic VTE not detected in daily practice, are assessed. Moreover, costs are yet increased by the treatment of these events usually not detected. Finally, because of a more rigorous follow-up, the patient's compliance is usually better in a RCT than in clinical practice, which has an impact on costs and effects of the studied treatment.

Blinding also has an impact on results of economic evaluations. Indeed, the fact that neither patients, nor providers have knowledge of the treatment group implies that all patients receive the same services and undergo the same tests, regardless of their treatment groups. Thus, artificial costs not present in clinical practice are induced by blinding.

Moreover, because of strict eligibility criteria, population in RCTs differs from population in the “real world”. Patients with co-morbidities or with an older age are for example often excluded in RCTs.

For the reasons exposed above, RCTs allow us to measure the efficacy and the theoretical cost of a treatment but not its efficiency, i.e. its costs and effectiveness in the daily practice. Moreover, they do not allow us to assess the quality of care in the daily practice.

2.1.1.2 Lack of relevant comparators

The standard procedure to correctly conduct economic evaluations is to compare an intervention with all relevant alternatives (17). However, in RCTs, drugs are frequently compared to placebo in order to deal with any placebo effect. So, except if no treatment is the standard of care, such strategy is not appropriate in economic evaluations. Moreover, when another treatment is used as comparator, its choice may not be appropriate or may be insufficient, other relevant treatments being not assessed. This factor, known by the name of “partial comparisons” is an important weakness of economic evaluations performed through RCTs.

2.1.1.3 Short time horizon

In RCTs, the time period captures a significant difference for specific outcomes of the treatment but not for all treatment effects and resources use during the natural history of the disease, as it is advised in economic evaluations. The study period in RCTs is thus often too short to identify all relevant outcomes. Surrogate markers such as the diminution of cholesterol in the blood are often measured but not final outcomes such as the life year saved or the quality-adjusted life-years (QALY) gained. Moreover, some adverse effects and their related costs will not be identified if they appear in the long term or if a patient is lost to follow-up because of complications.

2.1.1.4 Insufficient sample size

In RCT, the sample size is calculated in order to estimate the minimum number of patients required to detect a clinically important difference on the clinical primary endpoint for a given power and a statistical significance level (18). However, an RCT powered to detect clinical differences is often underpowered to detect differences in cost measures. Indeed, because of the high variability of health care resources use, economic evaluations require larger sample size than clinical trials. Moreover, some RCTs may be powered to detect differences for an intermediary outcome such as the reduction in blood pressure and not to detect differences in final outcomes such as the life year saved, which are required in economic evaluations.

2.1.1.5 Issues concerning ethical and cost aspects.

In some cases, RCTs are not feasible because of their costs or because of the ethical aspect of randomization. As example, for life-

threatening diseases, not providing the effective treatment to some patients because of randomization is not ethical.

To perform a multi-center RCT with a large sample size, a long period of follow-up, an important number of comparators and less strict eligibility criteria would be a solution to reduce the limitations cited above but such kind of study, known by the name of Practical Clinical Trial (19), is often too expensive compared to the benefits generated.

RCTs are thus not the best design to conduct an economic evaluation, especially when the primary outcome is an event with a low frequency of occurrence (+/- 1%) or when he implies a long term follow-up, as for chronic diseases (14). Moreover, if the main objective of the study is to analyze the care practices and the quality of care, other study designs are needed.

2.1.2 The contribution of observational and modelling studies

Observational and modelling studies are two well-known study designs overcoming the limitations of RCTs (14).

2.1.2.1 Observational studies

To measure the effectiveness of an intervention and the quality of care, observational studies are more appropriate than experimental studies because they are based on the observation of real world data.

Registries are potential sources of real world data and can be used to perform prospective cohort studies of patients with a specific disease and/or undergoing a specific intervention. Registries provide a larger

and more diverse population and reflect daily practices. So, they allow to overcome the main limitations of RCTs. However, in some situations as when the primary outcome has a low frequency of occurrence, the study must be relatively long and expensive in order to measure enough events. In such situation, retrospective observational studies performed from existing administrative databases are more appropriate (19).

Retrospective observational studies from administrative databases have a high external validity, are relatively inexpensive to use and are fairly quick to conduct because data were already collected. However, the major weaknesses are the selection bias due to the potential differences among the group studied and the lack of controls for confounding factors. Indeed, the control is only possible for confounding variables that are known and measured. Moreover, because data were collected for another purpose than the study, some important data like the presence of risk factors may not be reported (20). Others factors like the pharmacological treatment delivered to the patients are also often missing, which does not allow the comparison of treatments from administrative databases. In the literature, we found studies conducted from administrative databases to identify the number of symptomatic venous thromboembolism events after major orthopedic interventions in daily practice but the prophylaxis given to the patients could not be identified, which was a major weakness in these studies (21-23).

2.1.2.2 Modelling studies

To identify every effect and resource use for a specific intervention is not always possible from empirical or observational studies, especially for chronic diseases where a long-term period is required.

Modelling studies are thus needed to extrapolate the cascade of costs and effects resulting from the intervention (20). Moreover, modelling techniques allow to synthesize the multiple sources of evidence of efficiency from both RCTs and observational studies, to test various hypotheses, to compare all relevant alternatives, and to extrapolate results for other settings (24). Economic evaluations based on modelling studies are also relatively inexpensive and quick, and the uncertainty of the parameters (such as cost data) may be handled by sensitivity analysis.

The two best known methods are decision trees and Markov models (25). Decision trees are flow charts representing all possible options concerning a decision and their consequences over time. For each alternative option, paths representing the hierarchy of events following the option, their probabilities, and their costs are modeled. At the end of the path, outcomes may be intermediary (e.g. a symptom) or final (e.g. death) (3;20;25;26). Decision trees have the advantage to be simple to understand and easy to replicate for external people. However, they are only appropriate for simple problems with a fixed short time horizon, few recurrences of events and parameters which are not time-dependent (25).

On the other hand, for decisions implying recurrent events and/or events with an ongoing risk over time, as for chronic diseases, Markov models are preferred. Indeed, when the risk is ongoing, three important factors have to be taken into account (3;20;25;26). Firstly, some events may occur more than once. Secondly, the time at which an event will occur is uncertain and differences in the timing of the events have to be considered. Events occurring in the long term do not have the same impact than immediate events and thus, all costs and outcomes have to be discounted. Thirdly, some parameters are

time-dependent and their variations over time may be important to consider, as for example changing mortality rate according to the patient age. Because Markov models allow to easily deal with these factors, there are largely used for decision concerning chronic diseases (27-29).

Markov models are state-transition models representing the evolution of a cohort of patients over time and assuming that individuals are always in one of a finite number of health states, called “Markov states”. These health states may for example be defined according to the disease stages. Patients will progress from one state to another in a fixed unit of time, referred to as “Markov cycle”. For each Markov state, an annual transition probability towards the following state, a cost and a utility are referred (26;30). So, with this method, results can be extrapolated for a lifelong period and final outcomes such as the number of life-years saved or of quality-adjusted life-year (QALY) gained can be measured. A specific case of Markov processes where transition probabilities are fixed over time is called a Markov chain (31).

An important weakness of Markov models is its “memory-less” character. Indeed, in standard Markov models, the probability to evolve towards another state is assumed to be independent of the previous states of the patient and to only depend on its current state, which is not always a realistic assumption (20). This drawback can be overcome to some extend by increasing the number of health states but such a solution also increases quickly the complexity of the model. Furthermore, Markov models are usually complex and not transparent for external people, increasing the difficulty to judge the validity of the model and to determine if the assumptions are reasonable and appropriate to the study setting. So, Markov models

have to be transparent and external people must assess the quality of these models before to use them.

2.2 Reviews of existing economic evaluations

Systematic reviews of economic evaluations are useful sources of information for health care decision makers, especially for decisions concerning the reimbursement of interventions. Guidelines on the way to conduct and report a systematic review can be found in the literature (32). As highlighted in the previous chapter, the transferability of results to the country setting and the need for assessing the quality of economic evaluations were two major problems raised by health care decision makers (4). In this thesis, we decided to focus on the aspect of the quality assessment of the studies. Even if the interest on full economic evaluations has increased these last years, quality assessment processes are less developed than in the field of clinical studies. So, it is interesting to consider the experience in the field of clinical studies in order to make recommendations for the quality assessment of economic evaluations.

2.2.1 Quality assessment instruments in the field of clinical studies and meta-analyses

In the literature concerning clinical studies, it has been highlighted that quality assessment was an important aspect when performing a review of clinical studies. Such process is useful in many stages of a review. Firstly, in the selection of primary studies, a minimum of quality threshold can be determined, as for example to only include

RCTs. Further, quality assessment is also useful to appropriately interpret the findings, to explain the heterogeneity of results between studies, and to make recommendations for future researches (12).

Instruments assessing the quality of clinical studies are usually specific to a certain study design. So, in reviews with various designs (RCTs, observational studies…), multiple check-lists must be used. On the other hand, it should be noted that a previous study has shown that one single instrument combining checklists for various designs may also be considered (33).

Moreover, it has been shown that during the data analysis, a classification of the studies according to a quality score has an important impact. Indeed, numerous studies have highlighted that results may be opposite between studies with a high quality and studies of poor quality, and that estimates of effects may be biased with poor quality design (34;35).

Quality assessment giving a global quality score may also be useful to synthesize the data through meta-analyses, where results of the studies are pooled and weighted according to the quality of the studies. However, the instrument used to measure the quality has an impact on results of meta-analysis. Multiple scales exist and a study comparing 25 instruments for RCTs has shown that summary estimates differed according to the scale used (36). Thus, the instrument used has an important impact on the results of a systematic review of clinical studies.

Other discussions concern the method used to weight the items. A recent study has analyzed the effect of using different method of weightings for a quality assessment tool. They used an equal weight for each item but also different weightings such as subjective weightings defined by the author according to the importance he gave

to the items considered. As result, the studies were ranked differently according to the method used to determine the weightings. Moreover, when the results of studies with a high quality score and a low quality score were compared, the impact of the level of the studies' quality on results differed according to the method used (37).

So, more and more people advice to consider each component of methodological quality individually and to assess their impact on results rather than to use a global quality score (36).

Beside the instrument used, some people also discuss the way to measure the quality of the clinical studies. It is usually accepted to assess their quality by at least two unmasked reviewers in an independent way (12), even if it has been shown that blinding of reviewers to names and institutions of the authors and to journals had an impact on the assessment. With blinding, quality scores were lower and the interrater agreement was better than without blinding (38).

Finally, it should be noted that some instruments focus more on the quality of reporting than on the study methodology used (i.e. on internal validity, or on both internal and external validity) and that, even if it was often considered, a trial with a poor reporting is not necessarily a study of poor methodology (39).

2.2.2 Quality assessment of economic evaluations

Based on these discussions, different questions emerged when assessing the quality of economic evaluations. Are the instruments assessing the quality of economic evaluations appropriate for multiple designs? If economic evaluations were classified using a global quality score, would the classification differ according to the

instrument used? Is the use of a global quantitative quality score needed and what is the impact of weighting the criteria according to their relative importance?

3 OBJECTIVES AND OUTLINE OF THE THESIS

As it was explain in the first chapter, health care decision makers do not use enough economic evaluations in their decision process and improvements in this field are needed. The aim of the thesis was thus to investigate a number of methods in the field of economic evaluations which could be used in order to improve the information transmitted to Belgian health care decision makers. Such investigations allow us to define methodological recommendations but also recommendations on clinical practice.

Improvements could be done at two levels, i.e. at the level of primary economic evaluations or at the level of systematic reviews of economic evaluations.

At the level of primary economic evaluations, the previous chapters of the thesis have highlighted the need for real world data adapted to the Belgian setting and have shown that conversely to the field of clinical studies, RCTs were not the best design to conduct economic evaluations, especially for the assessment of outcomes with a low frequency of occurrence or for chronic diseases implying a long term follow-up.

To assess events with a low frequency of occurrence, the previous chapter raised the interest of using existing administrative databases

(see section 2.1.2). However, these databases are collected for another purpose than to perform economic evaluations and thus, do not include all important data. In particular, in the literature, we found studies performed from administrative databases in which information on drugs delivered to the patients were not included, which does not allow the assessment of the quality and the effectiveness of pharmacological treatments (22;40;41). On the other hand, Belgium has the originality to systematically collect information on reimbursed pharmacological treatments delivered to hospitalized patients. An association of this database with another routinely collected hospital database containing medical and administrative data, i.e. the Medical Minimum Basic Data Set (MBDS) database, is thus a potential source of information for Belgian health care decision makers. The aim of this part of the thesis was thus to determine if these routinely collected Belgian administrative databases contain enough information to assess the quality and the effectiveness of pharmacological treatments in the daily practice and thus to perform economic evaluations from real world data adapted to the Belgian setting.

Then, to analyze the treatment of chronic diseases, the previous chapter also highlighted that Markov models were especially appropriate (see section 2.1.2). In this part of the thesis, the objective was thus to construct a Markov model with parameters as much as possible specific to the Belgian situation and to investigate whether the analyzed treatment could become inefficient in daily practice.

Results and recommendations provided in these two parts are thus mainly specific to the Belgian setting.

At the level of systematic reviews of economic evaluations, the major concerns of health care decision makers were the quality of the economic evaluations and their transferability to the country setting. In this part of the thesis, we decided to focus on the quality assessment aspect. Based on the experience of clinical studies, the following research questions were investigated: Are the instruments assessing the quality of economic evaluations appropriate for multiple designs? If economic evaluations were classified using a global quality score, would the classification differ according to the instrument used? Is the use of a global quantitative quality score needed and what is the impact of weighting the criteria according to their relative importance? For this section, recommendations had thus an international scope.

To have a global vision of the outline of the thesis, figure 1 resumes the main sources of information in the field of economic evaluations and highlights, for each level, the methods examined in this thesis. It should be noted that in this figure, modelling studies include the area of both RCTs and observational studies. The choice of this graphical representation was based on the fact that modelling studies allow to combine these two sources of information.

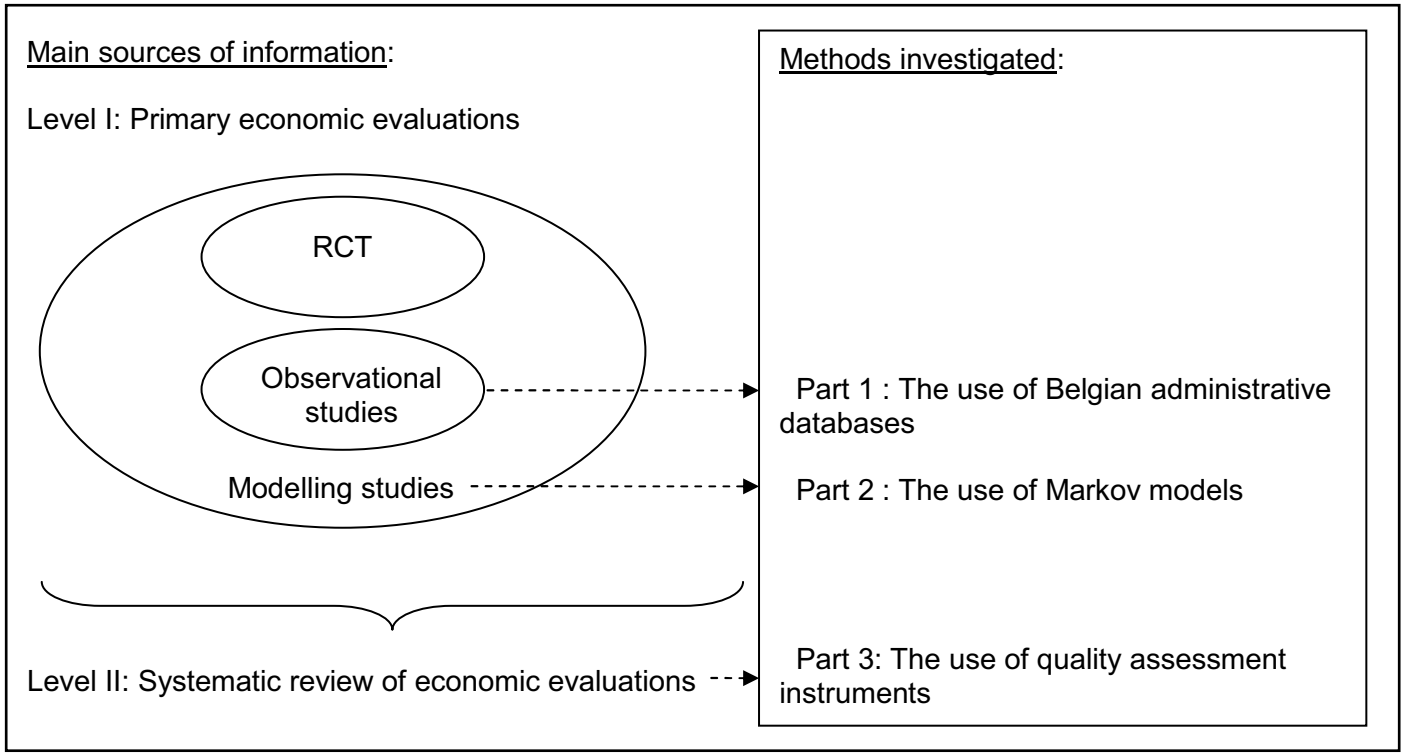

Figure 1 : Sources of information and methods investigated

The following chapters of this thesis are regrouped in three parts: the use of administrative databases, the use of Markov models and the use of quality assessment instruments. It should be noted that in the two first parts, a specific disease is analyzed in order to highlight the usefulness of the method for the characteristics of the disease. So, concerning the use of administrative databases, prophylaxis drugs to prevent venous thromboembolism after major orthopaedic surgeries are analyzed (i.e. events with low frequency (+/- 1%)). Then, concerning the use of Markov models, patients with chronic hepatitis C are analyzed (i.e. a chronic disease). Finally, in the third part, quality assessment instruments were used in a practical exercise on economic evaluations of the surgical treatment of obesity.

At the beginning of each part, the main research questions and justifications of methodological choices are exposed. Then, the performed studies on the use of these instruments are reported, followed by a conclusion giving recommendations to improve both the clinical practice and methodological aspects. Perspectives for future researches are also discussed. Finally, a global discussion summarizing the main findings and contributions of the thesis is presented.

References

(1) Dormont B, Grignon M, Huber H. Health expenditure growth: reassessing the threat of ageing. Health Econ 2006 Sep;15(9):947-63.

(2) Evans RG, Stoddart GL. Produire de la santé, consommer des soins. In: Evans RG, Barer ML, Marmor TR, editors. Être ou ne pas être en bonne santé: biologie et déterminants sociaux de la maladie.Montréal: Les Presses de l'Université de Montréal; 1996. p. 37-76.

(3) Drummond MF, Sculpher MJ, Torrance JW, O'Brien BJ, Stoddart JL. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford university press; 2005.

(4) Hoffmann C, Stoykova BA, Nixon J, Glanville JM, Misso K, Drummond MF. Do health-care decision makers find economic evaluations useful? The findings of focus group research in UK health authorities. Value Health 2002 Mar;5(2):71-8.

(5) Ofman JJ, Sullivan SD, Neumann PJ, Chiou CF, Henning JM, Wade SW, et al. Examining the value and quality of health economic analyses: implications of utilizing the QHES. J Manag Care Pharm 2003 Jan;9(1):53-61.

(6) Ament A, Evers S, Goossens M, De Vet H, Van Tulder M. Criteria list for conducting systematic reviews based on economic evaluation studies - the CHEC project. In: Donaldson C, Mugford M, Vale L, editors. Evidence-based Health Economics. From effectiveness to efficiency in systematic review.London: BMJ Books; 2002. p. 99-113.

(7) Jefferson T, Demicheli V, Vale L. Quality of systematic reviews of economic evaluations in health care. JAMA 2002 Jun 5;287(21):2809-12.

(8) Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. BMJ 1996 Aug 3;313(7052):275-83.

(9) Neumann PJ, Stone PW, Chapman RH, Sandberg EA, Bell CM. The quality of reporting in published cost-utility analyses, 1976-1997. Ann Intern Med 2000 Jun 20;132(12):964-72.

(10) Neumann PJ, Greenberg D, Olchanski NV, Stone PW, Rosen AB. Growth and quality of the cost-utility literature, 1976-2001. Value Health 2005 Jan;8(1):3-9.

(11) Woolf SH. Evidence-based medicine and practice guidelines: an overview. Cancer Control 2000 Jul;7(4):362-7.

(12) Khan KS, ter Riet G, Popay J, Nixon J, Kleijnen J. Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews. Stage II: Conducting the review. Phase 5: Study quality assessment. York: NHS Centre for Reviews and Dissemination, University of York; 2001 Mar. Report No.: CRD report 4 (2nd ed.).

(13) Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. J Clin Nurs 2003 Jan;12(1):77-84.

(14) Gagnon JP, Rindress D, Dix Smith M. Pharmacoeconomics: identifying the issues. Advisory panel reports. New Jersey: Internal Society for Pharmacoeconomics and Outcomes Research; 1998.

(15) O'Sullivan AK, Thompson D, Drummond MF. Collection of health-economic data alongside clinical trials: is there a future for piggyback evaluations? Value Health 2005 Jan;8(1):67-79.

(16) van Driel M. Glossaire des termes utilisés en Evidence-Based Medicine. Gent: Minerva; 2004.

(17) National Institute for Health and Clinical Excellence. The guidelines manual. Chapter 8: Incorporating health economics in guidelines and assessing resource impact. London: National Institute for Health and Clinical Excellence; 2007 Apr.

(18) Freedman KB, Bernstein J. Sample size and statistical power in clinical orthopaedic research. J Bone Joint Surg Am 1999 Oct;81(10):1454-60.

(19) Garrison LP, Neumann PJ, Erickson P, Marshall D, Mullins D. Using real world data for coverage and payment decisions: the ISPOR real world data task force report. New Jersey: International Society for Pharmacoeconomics and Outcomes Research.; 2006 Jul 25.

(20) Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.

(21) Oster G, Ollendorf DA, Vera-Llonch M, Hagiwara M, Berger A, Edelsberg J. Economic consequences of venous thromboembolism

following major orthopedic surgery. Ann Pharmacother 2004 Mar;38(3):377-82.

(22) White RH, Romano PS, Zhou H, Rodrigo J, Bargar W. Incidence and time course of thromboembolic outcomes following total hip or knee arthroplasty. Arch Intern Med 1998 Jul 27;158(14):1525-31.

(23) White RH, Zhou H, Romano PS. Incidence of symptomatic venous thromboembolism after different elective or urgent surgical procedures. Thromb Haemost 2003 Sep;90(3):446-55.

(24) Ades AE, Sculpher M, Sutton A, Abrams K, Cooper N, Welton N, et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. Pharmacoeconomics 2006;24(1):1-19.

(25) Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? Eur J Health Econ 2003;4:143-50.

(26) Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Making 1993 Oct;13(4):322-38.

(27) Plosker GL, Keam SJ. Candesartan cilexetil: a pharmacoeconomic review of its use in chronic heart failure and hypertension. Pharmacoeconomics 2006;24(12):1249-72.

(28) Shepherd J, Jones J, Hartwell D, Davidson P, Price A, Waugh N. Interferon alpha (pegylated and non-pegylated) and ribavirin for the treatment of mild chronic hepatitis C: a systematic review and economic evaluation. Health Technol Assess 2007 Mar;11(11):1-205, iii.

(29) Takeda A, Jones J, Shepherd J, Davidson P, Price A. A systematic review and economic evaluation of adefovir dipivoxil and pegylated interferon-alpha-2a for the treatment of chronic hepatitis B. J Viral Hepat 2007 Feb;14(2):75-88.

(30) Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics 1998 Apr;13(4):397-409.

(31) Chilcott J, Brennan A, Booth A, Karnon J, Tappenden P. The role of modelling in prioritising and planning clinical trials. Health Technol Assess 2003;7(23):iii, 1-iii125.

(32) NHS Centre for Reviews and Dissemination. Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews. York: NHS Centre for

Reviews and Dissemination, University of York; 2001 Mar. Report No.: CRD report 4 (2nd ed.).

(33) Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun;52(6):377-84.

(34) Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 1998 Aug 22;352(9128):609-13.

(35) Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, et al. Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. Health Technol Assess 1999;3(12):i-98.

(36) Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999 Sep 15;282(11):1054-60.

(37) Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med Res Methodol 2005;5:19.

(38) Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996 Feb;17(1):1-12.

(39) Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 2001 Jul 7;323(7303):42-6.

(40) Oster G, Ollendorf DA, Vera-Llonch M, Hagiwara M, Berger A, Edelsberg J. Economic consequences of venous thromboembolism following major orthopedic surgery. Ann Pharmacother 2004 Mar;38(3):377-82.

(41) White RH, Zhou H, Romano PS. Incidence of symptomatic venous thromboembolism after different elective or urgent surgical procedures. Thromb Haemost 2003 Sep;90(3):446-55.

PART 1 : THE USE OF ADMINISTRATIVE DATABASES

1 METHODOLOGICAL CHOICES AND RESEARCH QUESTIONS

Because administrative databases are especially useful to analyze events with a low frequency of occurrence (+/- 1%), we decided to assess the use of prophylaxis with low-molecular-weight heparin (LMWH) in the prevention of venous thromboembolism after major orthopedic surgery. The incidence of symptomatic venous thromboembolism event (VTE) after major orthopedic surgery has already been assessed in previous studies (1-3). However, in these studies, the prophylaxis given to the patients could not be identified and thus, their efficiency could not be analyzed.

The aim of this part was to determine if routinely collected Belgian administrative databases can be used to assess the quality of pharmacological treatments and more especially, the quality of low-molecular-weight heparin (LMWH) prophylaxis after major orthopedic surgery. Major orthopaedic procedures considered were total hip replacement (THR); total knee replacement (TKR) and hip fracture surgery (HFS).

Because we had a privileged access to data of one university hospital, the test was performed on this hospital.

Two administrative databases were analyzed in this hospital, i.e. the Medical Minimum Basic Data Set (MBDS) database containing

medical and administrative data of hospitalized patients and the health bill summary database containing information on reimbursed pharmacological treatments delivered to hospitalized patients.

Data about the incidence of events, the use of pharmacological treatment, and healthcare charges could be collected easily. Consequently, we analyzed the quality of care for prophylaxis with LMWH by assessing effectiveness criteria in terms of the incidence of symptomatic VTE, major bleeding and death in daily practice but also by evaluating adherence to guidelines and by investigating the costs of care.

Among effectiveness criteria assessed in this study, the major criterion was the incidence of symptomatic VTE. These symptomatic VTE events included thrombosis of superficial vein (SVT), deep-vein thrombosis (DVT) and pulmonary embolism (PE).

Because the incidence of these events was low (+/- 1%), specific statistical analyses must be considered. Indeed, the use of the maximum likelihood method is usually the preferred method in logistic regression, except under some conditions. For small or sparse datasets, or when the percentage of values equal to one in the datasets is close to 1%, the maximum likelihood method can fail or produce poor results. In these cases, exact logistic regressions are preferred (4). We thus analyzed these events using exact logistic regressions.

The list of ICD-9-CM codes used to identify patients undergoing THR, TKR, and HFS is described in table 1.

Table 1. List of ICD-9-CM codes

Major orthopedic procedure	ICD-9-CM	Description
THR	81.51	Total hip replacement - Replacement of both femoral head and acetabulum by prosthesis – Total reconstruction of hip
	81.53	Revision of hip replacement, not otherwise specified - Revision of hip replacement, not specified as to components(s) replaced, (acetabular, femoral or both)
TKR	81.54	Total knee replacement - Bicompartmental - Tricompartmental - Unicompartmental (hemijoint)
	81.55	Revision of knee replacement, not otherwise specified
HFS	81.52	Partial hip replacement - Bipolar endoprosthesis
	79.05	Closed reduction of fracture without internal fixation (site = femur)
	79.15	Closed reduction of fracture with internal fixation (site = femur)
	79.25	Open reduction of fracture without internal fixation (site = femur)
	79.35	Open reduction of fracture with internal fixation (site = femur)

Moreover, a list of all abbreviations used in this part can be found in table 2.

Table 2 : List of abbreviations

APR-DRG	All patient refined diagnosis related group
CI	Confidence interval
CNK	Code national(e) kode
DRG	Diagnosis related group
DVT	Deep-vein thrombosis
HFS	Hip fracture surgery
ICD-9-CM	International classification of diseases, 9th revision, clinical modification
LMWH	Low-molecular-weight heparin
LOS	Length of stay
MBDS	Minimum basic data set
MDC	Major diagnosis category
NSAIDs	Non-steroidal anti-inflammatory drugs
OR	Odds ratio
PE	Pulmonary embolism
RIZIV/INAMI	National health institute (Rijksinstituut voor ziekte- en invaliditeitsverzekering / Institut national d'assurance maladie-invalidité)
SVT	Thrombosis of superficial vein
THR	Total hip replacement
TKR	Total knee replacement
VTE	Venous thromboembolism

References

(1) Oster G, Ollendorf DA, Vera-Llonch M, Hagiwara M, Berger A, Edelsberg J. Economic consequences of venous thromboembolism following major orthopedic surgery. Ann Pharmacother 2004 Mar;38(3):377-82.

(2) White RH, Romano PS, Zhou H, Rodrigo J, Bargar W. Incidence and time course of thromboembolic outcomes following total hip or knee arthroplasty. Arch Intern Med 1998 Jul 27;158(14):1525-31.

(3) White RH, Zhou H, Romano PS. Incidence of symptomatic venous thromboembolism after different elective or urgent surgical procedures. Thromb Haemost 2003 Sep;90(3):446-55.

(4) King EN, Ryan TP. Statistical Practice : A preliminary investigation of maximum likelihood logistic regression versus exact logistic regression. American Statistical Association 2002 Aug;56(3):163-70.

2 ASSESSING THE QUALITY OF PHARMACOLOGICAL TREATMENTS FROM ADMINISTRATIVE DATABASES: THE CASE OF LOW-MOLECULAR-WEIGHT HEPARIN AFTER MAJOR ORTHOPEDIC SURGERY

S Gerkens, C Beguin, R Crott, MC Closon, Y Horsmans. Assessing the quality of pharmacological treatments from administrative databases: The case of low-molecular-weight heparin after major orthopedic surgery. Journal of Evaluation in Clinical Practice 2008 Aug;14(4): 585-94.

NB: The definitive version is available at www.blackwell-synergy.com.
© 2008 Blackwell Publishing Ltd. In this article, the terms “Dextran” instead of “Dextrose” (p588) and “effectiveness” instead of “effectivity” should have been used.

Journal of Evaluation in Clinical Practice ISSN 1356-1294

Assessing the quality of pharmacological treatments from administrative databases: the case of low-molecular-weight heparin after major orthopaedic surgery

Sophie Gerkens MSc,[1] Claire Beguin MD PhD,[2] Ralph Crott PhD,[3] Marie-Christine Closon PhD[4] and Yves Horsmans MD PhD[5]

[1]Researcher, [4]Professor, Université catholique de Louvain, School of Public Health, Brussels, Belgium
[2]Specialist, [3]Consultant, [5]Professor, Université catholique de Louvain, Cliniques universitaires Saint-Luc, Brussels, Belgium

Keywords
administrative databases, health economics, low-molecular-weight heparin, major orthopaedic surgery, quality of care

Correspondence
Sophie Gerkens
Université catholique de Louvain
School of Public Health
Unité de socioéconomie de la santé (SESA)
30 Clos Chapelle-aux-Champs box 3041
B-1200 Brussels
Belgium
E-mail: sophie.gerkens@infm.ucl.ac.be

Accepted for publication: 5 July 2007

doi:10.1111/j.1365-2753.2007.00926.x

Abstract

Rationale, aims and objectives 'Real world data' are needed to assess the quality of pharmacological treatments in clinical practice. The aim of this study was to determine whether administrative databases can be used to assess the quality of prophylaxis with low-molecular-weight heparin after major orthopaedic surgery.

Methods The study was performed in a Belgian university hospital. Patients undergoing total hip replacement (THR), total knee replacement (TKR) or hip fracture surgery (HFS) were selected retrospectively from the hospital's 2002 and 2003 administrative databases. Readmissions during the same year as the procedure were also analysed. Three quality indicators were assessed: incidence of venous thromboembolism (VTE), major bleeding and death; adherence to guidelines; and the costs of care.

Results Although 70% of data were collected from administrative databases, patients' records also had to be examined. During the period studied, VTE and major bleeding events were rare. Patients undergoing HFS were at greater risk of having a pulmonary embolism [Exact odds ratio (OR) = 3.78; 95% confidence interval (CI) = 1.13–16.22; $P = 0.03$] or of death from any cause (Exact OR = 2.15; 95% CI = 1.52–infinity; $P < 0.01$) than patients undergoing THR or TKR. The hospital's prophylaxis protocol was not always followed. Half the patients received higher prophylaxis doses than recommended and 11% received lower doses but no impact on adverse events was demonstrated.

Conclusion Results show that administrative databases contain enough information to measure the frequency of adverse events but complementary data on patient weight and on non-reimbursed drugs must be extracted from the patients' records to evaluate adherence to guidelines. Our findings stress the need for better integration of information systems.

Introduction

Randomized clinical trials are considered the gold standard to provide evidence on the efficacy of pharmacological treatments. However, in these studies, outcomes are measured in an 'ideal' and 'controlled' environment, which does not reflect daily practice. 'Real world' data analyses are needed to provide information on the quality of care [1] and on the effectivity of a treatment, i.e. its effect on daily practice. Administrative databases are potential sources of such information. These databases have the advantage of being readable on the computer, of encompassing large populations, and, once built, of being inexpensive to use [2–4]. They are mainly useful to study rare events [5]. Over the last decade, a comprehensive data-collecting system covering all in-hospital and day-case stays has been developed in Belgium. Routinely collected data from each hospital are centralized in national databases and are used for Public Health policy development and hospital funding. These databases have the originality to provide information on all reimbursed pharmacological products used during the in-hospital stay.

The aim of this study was to determine whether the standardized administrative databases of a Belgian university hospital ('Cliniques Universitaires Saint-Luc') could be used to assess the quality of pharmacological treatment in the hospital and to identify cost reduction opportunities. Such information would be interesting not only for medical care providers but also for health care purchasers and decision makers. We chose, as an example, the case of prophylaxis with low-molecular-weight heparin (LMWH) after major

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd,

orthopaedic surgery to reduce the incidence of venous thromboembolism (VTE). The studied hospital has 960 beds, with 90 beds for the orthopaedic unit. Because VTE complications after major orthopaedic surgery are mostly managed in hospital, quality indicators including readmissions for VTE events could be evaluated.

The Institute of Medicine has defined quality of care as 'the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge' [6]. An important aspect of quality of care is consistency with guidelines. Because data about the incidence of events, use of pharmacological treatment and health care charge could be collected easily from the studied hospital, we analysed the quality of care for prophylaxis with LMWH not only by assessing effectivity criteria in terms of the incidence of symptomatic VTE, major bleeding and death in daily practice but also by evaluating adherence to guidelines and investigating the costs of care.

Methods

Description of databases

Two standardized administrative databases are developed: the Medical Minimum Basic Data Set (MBDS) database and the health bill summary database. The Medical MBDS database contains medical information, such as principal diagnosis, secondary diagnoses and procedures, from which the Major Diagnosis Category, All Patient Refined Diagnosis Related Group (APR-DRG) and a measure of the severity of the patient's hospitalization can be derived. Administrative data, such as length of stay, status at discharge (including death) and demographic information (such as age and gender) are also included. Diagnoses and procedures are coded using the International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM).

The health bill summary database contains information on resource use, such as the types and prices of performed procedures and of all pharmacological products given to the patient. Reimbursed pharmacological products are identified by their National Health Insurance Institute (RIZIV/INAMI) codes and non-reimbursed products are grouped under a single code.

Selection of in-hospital stays

An observational retrospective study of the university hospital's 2002 and 2003 databases was performed. The most frequent major orthopaedic procedures were studied, i.e. total hip replacement (THR), total knee replacement (TKR) and hip fracture surgery (HFS) [7]. These in-hospital stays were selected from the Medical MBDS database using the ICD-9-CM procedure codes (THR: 81.52–81.53; TKR: 81.54–81.55; HFS: 81.52–79.05–79.15–79.25–79.35).

Quality indicators

Three quality indicators were assessed: effectivity criteria, adherence to guidelines and costs.

Effectivity criteria

To measure effectivity criteria, deaths, symptomatic VTE and adverse effects were analysed. In the literature, adverse effects of LMWH are mostly identified as major bleeding and thrombocytopenia [7–9], but the definition of major bleeding varies between studies [7,10–12]. In this study, major bleeding was defined as bleeding in a critical organ, i.e. retroperitoneal bleeding, cerebral bleeding, ocular bleeding, gastrointestinal bleeding or urinary bleeding. Symptomatic VTE and adverse effects were identified from the Medical MBDS database using the ICD-9-CM diagnosis codes. For thrombocytopenia, no event was observed as being due to heparin in this study. This adverse effect was therefore not analysed. The number of deaths was also assessed from the National Medical MBDS database, but to determine the cause of death, patients' records had to be investigated.

Because a great number of VTE appeared in the 3 months after discharge [13], an analysis of events after discharge was also performed. However, data about outpatient visits or readmissions in other hospitals were not available. Consequently, the number of events after discharge in this study was only measured by the number of readmissions for VTE or major bleeding in the same hospital. The period taken into account was the same year as the major orthopaedic surgery.

Adherence to guidelines

To evaluate the adherence to guidelines, the hospital's protocol was first analysed. In 2002 and 2003, two prophylaxis drugs were recommended in this hospital, nadroparin and enoxaparin. For nadroparin, a dose of 0.3 mL day^{-1} during the post-operative period was recommended for patients weighing less than 70 kg, 0.4 mL day^{-1} for patients weighing between 70 and 90 kg, 0.6 mL day^{-1} for patients weighing between 91 and 120 kg, and 0.8 mL day^{-1} for patients weighing more than 120 kg. For enoxaparin, a dose of 40 mg day^{-1} during the post-operative period was recommended for patients weighing less than 100 kg, and 60 mg day^{-1} for patients weighing more than 100 kg. No recommendation was specified for the preoperative period and the daily dose delivered in this period depended on individual clinical choices.

The types and doses of prophylaxis given were then analysed. Reimbursed prophylaxis drugs were detected from the health bill summary database using INAMI codes. However, neither the Medical MBDS database nor the health bill summary database could be used to identify non-reimbursed prophylaxis. Moreover, for LMWH, the daily dose to be delivered was dependent on patient weight and this variable was also not available in the administrative databases. Additional sources of data were thus required. Patients' hospital records were used to identify patient weight and the hospital's internal billing data were analysed to determine the type and dose of any non-reimbursed prophylaxis.

Finally, subgroups were created according to protocol adherence. As explained in Table 1, three subgroups were firstly determined from internal billing data: people who had received 'lower' doses than recommended, patients who had received the 'correct' doses, and patients who had received 'higher' doses than recommended in the protocol. Because no general rule existed for the preoperative period, a fourth subgroup called 'possibly higher' was also created. Patients who had received prophylaxis other than nadroparin or enoxaparin, patients who had changed prophylaxis during the stay, and patients for whom no prophylaxis treatment

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

Table 1 Subgroup determination for protocol adherence

Subgroup	Characteristics
Correct dose	Protocol dose day^{-1} × post-operative time < x < protocol dose day^{-1} × (post-operative time + p + 4*)
Lower dose than recommended	x < protocol dose day^{-1} × post-operative time
Higher dose than recommended	x > protocol dose day^{-1} × (length of the in-hospital stay + 4*)
Possibly higher dose than recommended	Protocol dose day^{-1} × (post-operative time + p + 4*) < x < protocol dose day^{-1} × (length of the in-hospital stay + 4*)
Other	Prophylaxis not included in the protocol, change of prophylaxis during the hospitalization, or no prophylaxis identified

x, total prophylaxis dose delivered to the patient recorded in the billing data; p, if the preoperative time = 0, then p = 0, otherwise p = 1.
*Four days were added to take into account the maximal prophylaxis dose patients could receive to take home.

Table 2 Genetic risk factors for VTE [14–16]

Risk factors	Sources
Deficiency of anti-thrombin III, deficiency of protein C or S, Factor V Leiden mutation, activated protein C resistance, prothrombin mutation (20210 A), hyperhomocystinaemia	MBDS database
Factor XII deficiency, dysfibrinogenaemia	MBDS and patients' records
Elevated levels of factor VIII, factor IX or factor XI, dysplasminogenaemia	Not identified

VTE, venous thromboembolism; MBDS, National Medical Minimum Basic Data Set.

was identified were then classified in a fifth subgroup, 'Others'. In the hospital, pharmacists issued the drugs two times per week; it was thus possible that patients received prophylaxis for a maximum of four supplementary days at discharge. These possible additional days of prophylaxis were taken into account in the subgroup calculation (see Table 1). Among patients with higher than recommended prophylaxis doses ($n = 545$), patients with atrial fibrillation, atrial flutter or patients receiving long-term anticoagulant therapy before the hospitalization ($n = 43$) were excluded from the 'higher' subgroup and placed in the 'possibly higher' subgroup, as these factors could have explained the higher LMWH doses delivered. When patients developed VTE or major bleeding events, electronic billing data could not be used, as it was not possible to distinguish prophylaxis doses given before the event from the doses delivered to cure the event. However, because these events were not frequent, patient records and the billing data for these in-hospital stays were searched manually to determine the prophylaxis doses given before the event and thus to determine the correct subgroups for these stays.

Cost analysis

Costs were calculated using official national tariffs. We simulated not only the direct health care charges from the perspective of the RIZIV/INAMI but also the direct health care related out-of-pocket expenses of patients. Possible savings in terms of increased productivity were not assessed. The daily price for the university hospital stay was €450. Costs related to VTE or major bleeding events (including the additional cost related to these events during the in-hospital stay for THR, TKR or HFS, and the costs of readmission for the event) and the additional cost related to higher prophylaxis doses than recommended were investigated.

The additional costs related to the development of VTE or major bleeding during the in-hospital stays for THR, TKR and HFS were calculated in three stages. For each stay, a DRG and a severity level were assigned. For stays with the observed events, the DRG and severity level were recalculated after exclusion of the event. This new DRG severity level allowed us to compare the observed costs of the stay with the cost if the event had not occurred and so to determine the additional cost linked to the event.

The cost of readmission was assessed by the total cost of readmission for VTE or major bleeding events in the hospital. A distinction was made between readmissions for which VTE or major bleeding events were the reason for the hospitalization (= events coded as principal diagnosis) and readmissions for other reasons (= events coded as secondary diagnosis).

For patients without VTE, major bleeding or death events, and who had received higher than recommended doses of prophylaxis, an excess prophylaxis cost was also estimated. First, for each of these patients, an expected prophylaxis cost was assessed by determining the correct dose the patient would have received if the protocol had been respected. Second, an excess prophylaxis cost per patient was obtained by subtracting the expected prophylaxis cost from the real prophylaxis cost in the billing data. A total excess prophylaxis cost could thus be obtained.

Risk factor determination

Lists of risk factors for VTE and major bleeding were constructed from data available in the current literature [14–16] (Tables 2–4). It should be noticed that patient's age and obesity (body mass index > 30), generally considered as risk factors in the literature [14–16], were identified separately for the purposes of this study because most patients undergoing major orthopaedic interventions are more than 40 years old and the prophylaxis dose delivered takes into account the risk linked to body weight. Reimbursed drugs that were noted as potential risk factors were identified from the health bill summary database

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

Table 3 Acquired risk factors for VTE [14–16]

Risk factors	Sources
Immobility (bed rest > 4 days), age > 40*, pregnancy, previous VTE, malignancy, phospholipid antibody or lupus anticoagulant, heart and respiratory failure, recent myocardial infarction (8 weeks), inflammatory bowel disease, nephrotic syndrome, polycythemia, paraproteinaemia, Behçet's disease, paroxysmal nocturnal haemoglobinuria, myeloproliferative diseases, acute myelogenous leukaemia	MBDS database
Obesity (BMI > 30)*	MBDS and patients' records
Puerperium	Not identified
Hormonal therapy	Hospital's billing data

*In this study, age and obesity were not considered risk factors.
VTE, venous thromboembolism; MBDS, National Medical Minimum Basic Data Set; BMI, body mass index.

Table 4 Risk factors for bleeding [14–16]

Risk factors	Sources
Defective haemostasis, hepatic dysfunction, renal failure, ulcer history, diabetic retinopathy, retinal vascular disease, angiodysplasia of stomach, angiodysplasia of intestine	MBDS database
Alcoholism	MBDS and patients' records
Cachexy (BMI < 15)	Patients' records
Use of NSAIDs or platelet aggregation inhibiting drugs during the stay	Hospital's billing data
Uncontrolled hypertension, use of concurrent treatment* during the stay, use of anticoagulant or anti-platelet drugs the week before hospitalization	Not identified

*Concurrent treatment = vitamin K antagonists, dextrose or thrombolytic drugs.
MBDS, National Medical Minimum Basic Data Set; NSAIDs, non steroidal anti-inflammatory drugs; BMI, body mass index.

Table 5 Relevant APR-DRG

APR-DRG	Description
300	Bilateral and multiple major joint procs of lower extremity
301	Major joint and limb reattach proc of lower extremity for trauma
302	Major joint and limb reattach proc of lower extremity except for trauma
308	Hip and femur procedures except major joint for trauma
309	Hip and femur procedures except major joint for non-trauma

APR-DRG, All Patient Refined Diagnosis Related Group.

using INAMI codes. For non-reimbursed drugs, this information was retrieved from the hospital's internal billing data. The use of anticoagulant or anti-platelet drugs in the week before the hospitalization could not be detected but, in this situation, the surgical procedure is reported. It was also not possible to assess whether vitamin K antagonists, dextran or thrombolytics were taken simultaneously with the LMWH treatment but, in practice, these drugs are usually stopped when LMWH prophylaxis starts.

Diagnoses noted as risk factors were identified from the Medical MBDS database using ICD-9-CM diagnoses codes. However, specific pathologies, such as elevated levels of factor VIII, IX, XI, dysplasminogenaemia, puerperium and uncontrolled hypertension could not be precisely identified by ICD-9-CM codes. These risk factors could therefore not be investigated. The ICD-9-CM diagnosis codes to detect patients with factor XII deficiency and dysfibrinogenaemia were ambiguous and also regrouped diseases with a risk of bleeding. For in-hospital stays with this diagnosis code, the patients' records were analysed in order to determine if these patients really were at risk of VTE.

Exclusion criteria

In-hospital stays for THR, TKR or HFS during the years 2002 and 2003 were retrospectively selected according to the ICD-9-CM procedure codes ($n = 1235$). Patients who were younger than 18 years of age ($n = 20$ stays) were excluded. To have a homogeneous population, only stays classified in APR-DRGs for hip or knee prostheses or surgical treatments of hip fracture (see Table 5) were retained ($n = 69$ stays excluded). For one stay, the patient had developed a pulmonary embolism (PE) before the intervention and received therapeutic doses of LMWH rather than prophylaxis doses. This stay was thus excluded from the analysis ($n = 1$ stay). Stays of patients with no weight mentioned in their records ($n = 46$ stays) were also excluded. During the period studied, some patients underwent multiple in-hospital stays for major orthopaedic surgery. For these patients, we decided to analyse only their first stay and excluded the others ($n = 74$ stays). For example, if a patient underwent THR in the beginning of the year and TKR at the end of the year, we only analysed the in-hospital stay for the THR. It should be noticed that for these subsequent excluded

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

hospitalizations, there were no VTE events, adverse effects or deaths. A total of 1025 in-hospital stays for major orthopaedic surgery were thus selected for this analysis.

Statistical analysis

Statistical tests were interpreted at a significance level of 0.05 and *P*-values were for a two-sided hypothesis test. Logarithmic transformation was used in case of non-normal distribution. For univariate analyses concerning patient and in-hospital stay characteristics, differences in the means of continuous variables were evaluated using one-factor variance analyses in the case of (log-) normal distribution, and a Wilcoxon rank sum test otherwise. Differences in proportions of dichotomous categorical variables were analysed with either chi-squared tests or Fisher exact tests, depending on the expected cell-size. For other categorical variables, chi-squared tests were used. To analyse the contribution of different factors to protocol adherence, a logistic regression analysis was performed with receiving higher than recommended doses of prophylaxis as the dependent variable. Because after major orthopaedic surgery, death, VTE and major bleeding events occur very infrequently, exact logistic regressions were used for univariate and multivariate analyses of these events, using the hybrid network and Monte Carlo algorithm of Mehta *et al*. [17,18]. The exact method led to larger confidence intervals (CIs) and in two cases to an upper limit of the CI tending towards infinity, implying that the effect of the variable was certain but badly defined. Tests were performed using SAS software V9. In all statistical analyses, the severity level of the in-hospital stays was recalculated after exclusion of the ICD-9-CM diagnoses codes for VTE or major bleeding. It should also be noticed that the variable 'risk of VTE' used in the analyses was characterized by the presence of a VTE risk factor other than the risk due to the major orthopaedic procedure itself.

Results

Patients' characteristics and description of in-hospital stays

The study population had a mean age of 70.3 years at the time of the major orthopaedic surgery, 68.6% were women, 18.5% were at risk of VTE, and 76.6% were at risk of bleeding, mostly because of the use of non-steroidal anti-inflammatory drugs (NSAIDs) (69.6%). There were significant differences in patient characteristics among groups. In the HFS group, the mean age, the proportion of women and the frequency of patients at risk of VTE were higher than that in the other groups, while the frequency of patients at risk of bleeding was lower. In the TKR group, the frequency of obese patients was higher than that in the other groups (Table 6).

The characteristics of the in-hospital stays were also significantly different among groups. In the HFS group, the mean preoperative time was shorter, the mean length of stay was longer, the frequency of emergency admissions was greater, and the frequency of stays with an extreme severity level was higher than that in the other groups (Table 6).

The incidence of symptomatic VTE events and adverse effects

Few patients developed symptomatic VTE (2.1%). The number of patients with PE was significantly different among the THR, TKR and HFS groups, with a higher proportion in the HFS group [Exact odds ratio $(OR)_{HFS\ -\ THR\ or\ TKR} = 3.78$; 95% CI = 1.13–16.22; $P = 0.03$]. This was mostly due to a higher proportion of readmissions for PE after HFS (Exact $OR_{HFS\ -\ THR\ or\ TKR} = 5.98$; 95% CI = 1.12–52.24; $P = 0.03$). Nine patients developed VTE events on readmission and for five of these patients it was the reason for the hospitalization. No patient was readmitted for VTE after TKR.

Table 6 Characteristics of patients and of in-hospital stays

	THR *n* = 422	TKR *n* = 271	HFS *n* = 332	*P*-value	Overall *n* = 1025
Mean age; years (range)	66.4 (18–95)	67.9 (18–89)	77.1 (18–101)	<0.0001	70.3 (18–101)
Gender (male/female); *n*/*n* (%/%)	153/269 (36.3/63.7)	83/188 (30.6/69.4)	86/246 (25.9/74.1)	0.0093	322/703 (31.4/68.6)
Obese patients (BMI > 30); *n* (%)	77 (18.2)	105 (38.7)	23 (6.9)	<0.0001	205 (20.0)
Patients at risk of VTE; *n* (%)	76 (18.0)	39 (14.4)	75 (22.6)	0.0338	190 (18.5)
Patients at risk of bleeding	339 (80.3)	236 (87.1)	210 (63.3)	<0.0001	785 (76.6)
NSAID use	320 (75.8)	226 (83.4)	167 (50.3)	<0.0001	713 (69.6)
Mean preoperative stay (range); days	2.3 (0–56)	1.7 (0–46)	1.3 (0–42)	<0.0001	1.8 (0–56)
Mean length of stay (range); days	17.5 (7–146)	16.2 (3–84)	23.6 (1–208)	<0.0001	19.1 (1–208)
Emergency; *n* (%)	57 (13.5)	4 (1.5)	319 (96.1)	<0.0001	380 (37.1)
Severity at admission; *n* (%)					
Mild	135 (32.0)	109 (40.2)	100 (30.1)	<0.0001	344 (33.6)
Moderate	217 (51.4)	147 (54.2)	140 (42.2)		504 (49.2)
Severe	68 (16.1)	13 (4.8)	76 (22.9)		157 (15.3)
Extreme	2 (0.5)	2 (0.7)	16 (4.8)		20 (2.0)

THR, total hip replacement; TKR, total knee replacement; HFS, hip fracture surgery; BMI, body mass index; VTE, venous thromboembolism; NSAID, non-steroidal anti-inflammatory drug.

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

The number of major bleeding events was also small and the difference among procedures was not significant (Exact $OR_{HFS - THR\ or\ TKR} = 5.32$; 95% CI = 0.85–54.87; $P = 0.08$). All patients who suffered from major bleeding were at risk of bleeding. Three patients had major bleeding on readmission and for two of these patients it was the reason for the hospitalization. The total number of deaths from any cause was significantly different among groups, with a higher proportion in the HFS group (Exact $OR_{HFS - THR\ or\ TKR} = 2.15$; 95% CI = 1.52–infinity; $P < 0.0001$). Among deceased patients, two had developed PE in the HFS group and one had suffered from major bleeding (Table 7).

We tried to explain the effects of different variables on the proportion of VTE events, major bleeding events and deaths during the hospitalization or on readmission. Because emergency admissions were strongly correlated with HFS procedures, this variable was not tested in the model (Spearman correlation coefficient: 0.846; $P < 0.0001$). For VTE events, the model included age, gender, obesity, the risk of VTE, the type of procedure, the severity level of the stay and the administration of lower than recommended prophylaxis doses. Only the severity level (severe or extreme levels = 1) showed a significant impact on the proportion of VTE events (Table 8). For major bleeding, the following variables were introduced in the model: age, gender, obesity, the type of procedure, the severity level of the stay, the administration of higher than recommended prophylaxis doses, NSAID use and the interaction between higher than recommended prophylaxis doses and NSAIDs. Only age had a significant impact on the proportion of major bleeding events (Table 8). Finally, we analysed the effect of age, gender, obesity, the risk of VTE, the risk of major bleeding, the type of procedure, the severity level of the stay, the administration of lower than recommended prophylaxis doses and the development of PE or major bleeding on the probability of death. HFS procedures, patient age and a severe or extreme level of severity had a significant impact on deaths from any cause (Table 8).

Adherence to the protocol

During the in-hospital stay, most patients were treated with nadroparin (94.1%). For a small number of patients (5.9%), no prophylaxis was identified from the billing data, another treatment than the prophylaxis advised in the protocol was administered, or the prophylaxis had changed during the hospitalization.

Half the patients (49%) received higher than recommended prophylaxis doses during the in-hospital stay and 11% received lower than recommended doses (Table 9).

We analysed the effect of age, gender, obesity, emergency admission, the severity level of the stay and the risk of VTE on the probability of receiving higher than recommended doses of prophylaxis during the hospitalization. Emergency admission, gender, obesity, the severity level of the stay and the risk of VTE had a significant impact on the probability of receiving higher than recommended doses of prophylaxis (Table 10).

Only one patient who received lower than recommended prophylaxis doses during hospitalization developed symptomatic VTE on readmission (Exact $OR_{inferior - normal\ or\ superior} = 0.89$; 95% CI = 0.02–7.09; $P = 1.00$). The probability of having a major bleeding event during the hospitalization or a readmission was not significantly different for patients who received higher

Table 7 Incidence of VTE, major bleeding and death events

	THR ($n = 422$)		TKR ($n = 271$)		HFS ($n = 332$)		Exact *P*-value*			Overall ($n = 1025$)
	In-hospital stay	After discharge	In-hospital stay	After discharge	In-hospital stay	After discharge	After discharge	In-hospital stay	Total	
Symptomatic VTE	4 (0.9)	3 (0.7)	5 (1.8)	0	4 (1.2)	6 (1.8)	1.00	0.07	0.27	22 (2.1)
Total SVT	1 (0.2)	2 (0.5)	1 (0.4)	0	1 (0.3)	0	1.00	0.92	0.96	5 (0.5)
Total DVT	1 (0.2)	1 (0.2)	3 (1.1)	0	2 (0.6)	3 (0.9)	1.00	0.21	0.40	10 (1.0)
Total PE	2 (0.5)	2 (0.5)	1 (0.4)	0	3 (0.9)	6 (1.8)	0.60	0.03	0.03	14 (1.4)
Major bleeding	1 (0.2)	0	0	1 (0.4)	3 (0.9)	2 (0.6)	0.20	0.51	0.08	7 (0.7)
Death from any cause	1 (0.2)	2 (0.5)	1 (0.4)	1 (0.4)	19 (5.7)	5 (1.5)	<0.0001	0.15	<0.0001	29 (2.8)
Death after PE	0	0	0	0	1 (0.3)	1 (0.3)	0.64	0.64	0.21	2 (0.2)
Death after major bleeding	0	0	0	0	0	1 (0.3)	NA	0.64	0.65	1 (0.1)

*From exact logistic regression with HFS as dependent variable.
VTE, venous thromboembolism; THR, total hip replacement; TKR, total knee replacement; HFS, hip fracture surgery; SVT, thrombosis of superficial vein; DVT, deep-vein thrombosis; PE, pulmonary embolism; NA, not appropriate.

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

Table 8 Multivariate analyses of factors associated with events

Dependent variables	Independent variables	Exact odds ratios	95% CI	*P*-value
Symptomatic VTE	Extreme or severe level of severity	3.46	1.29–11.89	0.013
Major bleeding	Age	1.11	1.02–1.25	0.013
Death from any cause	HFS	4.09	1.70–14.94	0.001
	Extreme or severe level of severity	2.04*	1.50–infinity	<0.0001
	Age	1.04	1.01–1.08	0.012

*Indicates a median unbiased estimate including the observed statistic.
CI, confidence interval; VTE, venous thromboembolism; HFS, hip fracture surgery.

Table 9 Characteristics of prophylaxis

	THR *n* = 422	TKR *n* = 271	HFS *n* = 332	Overall *n* = 1025
Type of prophylaxis*; *n* (%)				
Nadroparin	396 (93.8)	253 (93.4)	316 (95.2)	965 (94.1)
Other	26 (6.2)	18 (6.6)	16 (4.8)	60 (5.9)
Subgroup repartition†				
Lower	45 (10.7)	32 (11.8)	34 (10.2)	111 (10.8)
Correct	116 (27.5)	99 (36.5)	81 (24.4)	296 (28.9)
Higher	212 (50.2)	115 (42.4)	175 (52.7)	502 (49.0)
Possibly higher	23 (5.5)	7 (2.6)	26 (7.8)	56 (5.5)
Other	26 (6.2)	18 (6.6)	16 (4.8)	60 (5.9)

*Enoxaparin was not administered.
†Representative of adherence to the protocol (see Table 3).
THR, total hip replacement; TKR, total knee replacement; HFS, hip fracture surgery.

Table 10 Multivariate analysis of factors affecting adherence to the protocol

Dependent variable*	Independent variables	Odds ratios	95% CI	*P*-value
Higher than recommended dose of prophylaxis	Emergency	1.37	1.04–1.81	0.0256
	Obesity	0.63	0.46–0.88	0.0061
	Gender (woman = 1)	1.45	1.11–1.91	0.0068
	Severity level (severe or extreme = 1)	0.57	0.40–0.82	0.0025
	To be at risk of VTE	1.46	1.05–2.05	0.0268

*A model with lower than recommended dose of prophylaxis as the dependent variable was also analysed but no variable had a significant impact.
CI, confidence interval, VTE, venous thromboembolism.

than recommended doses than for patients with correct or lower than recommended doses (Exact OR = 0.62; 95% CI = 0.09–3.69; $P = 0.80$). The number of deaths for any cause during the hospitalization was significantly different for patients who received lower than recommended doses than for patients with correct or higher than recommended doses (Exact OR = 5.67; 95% CI = 1.41–40.67; $P = 0.01$) but the number of deaths after VTE or major bleeding events was not significantly different between these subgroups. Moreover, in the multivariate analysis, receiving lower than recommended doses of prophylaxis had no significant impact on the probability of death from any cause. Thus, although the protocol was not followed for more than half of patients, this had no significant impact on events in this study.

Cost analysis

Without any events, mean and median charges of in-hospital stays ranged from €5662 to €52 598 per APR-DRG and severity level. Of the 1025 patients, 1.3% developed VTE events and 0.4% developed major bleeding during in-hospital stays for THR, TKR or HFS. VTE events increased the cost of an in-hospital stay by an average of 27%. The mean additional cost due to VTE was €3920 per patient and the median additional cost was €5505 per patient. The mean additional cost due to major bleeding was €19 265 per patient (+125%) and the median additional cost was €12 191 per patient (+86%) (Table 11). The mean costs of readmission for VTE and for major bleeding as a principal diagnosis were €7684 and €3132 respectively (Table 12).

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

Table 11 Additional cost due to events during in-hospital stays for THR, TKR or HFS

		Δ LoS		Total additional cost (increase in percentage)	
	n (%)	Mean	Median	Mean	Median
VTE	13 (1.3)	10	9	€3919.9 (+26.8%)	€5504.9 (+43.9%)
Major bleeding	4 (0.4)	36	19	€19 264.7 (+125.2%)	€12 191.2 (+85.8%)

THR, total hip replacement; TKR, total knee replacement; HFS, hip fracture surgery; Δ, differential due to the event; LoS, length of stay; VTE, venous thromboembolism.

Table 12 Cost of events after discharge

	Principal diagnosis				Secondary diagnosis			
	Mean		Mean total		Mean		Mean total	
	n (%)	LoS	Cost	Time	*n* (%)	LoS	Cost	Time
VTE	5 (0.5)	14	€7683.7	<3 months	4 (0.4)	18	€9833.3	≤3 months*
Major bleeding	2 (0.2)	6	€3132	3–7 months	1 (0.1)	6	€10 748.6	<3 months

*Except for one patient for whom the event occurred after 6 months.
LoS, length of stay; VTE, venous thromboembolism.

For patients without VTE, major bleeding or death who received higher than recommended doses of prophylaxis ($n = 492$), the total excess prophylaxis cost was €8380, i.e. on average €17 per patient. Among these patients, the mean excess prophylaxis cost for patients with VTE risk factors was €27 per patient.

Discussion

As expected, our study shows that patients undergoing HFS had a greater probability of developing PE or of death from any cause than patients undergoing THR or TKR. This group concentrates patients at higher risk, i.e. patients admitted as an emergency (96.1% of patients in the HFS group compared with 13.5% in the THR group and 1.5% in the TKR group). This finding has also been observed in randomized double-blind trials where the incidence of death events by day 11 with enoxaparin as prophylaxis was also higher in HFS than that in THR and TKR [19–22]. However, VTE and death events after HFS were higher in our study than in these randomized trials [19–22]. The study designs, especially the exclusion criteria of the randomized trials, could explain these apparent differences. On the other hand, comparisons of our results with studies performed from administrative databases [23–25] also show a higher rate of in-hospital symptomatic VTE, mostly for TKR and HFS. An explanation for this finding could be the shorter lengths of stay and the higher rates of symptomatic VTE after discharge reported in these studies [23–25]. However, these studies did not report the prophylaxis used or its impact on VTE events.

The aim of our study was to determine whether administrative databases routinely collected by the Ministry of Health contain enough information to assess the quality of care related to in-hospital pharmacological treatment, i.e. LMWH prophylaxis after major orthopaedic surgery. In this case, 70% of the data needed could be extracted from the Medical MBDS and from the health bill summary databases, allowing us to assess the incidence and cost of VTE events and the adverse effects of LMWH. These results could then be compared with those of previous studies. They could also be used as indicators for quality of care in order to establish benchmarks and to make comparisons among Belgian hospitals. Nevertheless, administrative databases present some limitations and do not allow adherence to guidelines or prophylaxis costs to be estimated. Indeed, estimation of the adherence to the protocol defined in the hospital required precise information on patient weight and on the prophylaxis administered. However, neither patient weight nor data on non-reimbursed drugs with an impact on prophylaxis could be collected from the standardized administrative databases. These complementary data must be extracted from the patients' records and from the hospital billing system.

This exercise in our hospital reveals some interesting findings. In most cases, the protocol recommended by the hospital does not seem to have been followed, i.e. half of the patients studied received higher than recommended doses of prophylaxis and 11% received lower than recommended doses. Nevertheless, among patients who received lower than recommended prophylaxis doses, only one developed symptomatic VTE and patients who received higher than recommended prophylaxis doses did not develop significantly more major bleeding events than patients with correct or lower than recommended doses.

Woman, non-obese patients, emergency admissions, the severity level of the stay and the risk of VTE had a significant impact on the probability of receiving higher than recommended doses of prophylaxis. These results could be interpreted as the result of physicians being afraid to give insufficient prophylaxis doses to these patients and/or as a tendency to give a standard dose to the patients, without taking weight into account. Physicians did not seem to use the lower prophylaxis dose (i.e. 0.3 mL day^{-1}), even if this was advised by the protocol. Consequently, patients with a low weight and thus recommended for low prophylaxis doses in the protocol were more likely to receive higher than recommended prophylaxis doses. Conversely, overweight patients had less chance, which could explain the lower probability of obese patients to receive superior doses of prophylaxis.

Seventy per cent of patients received NSAIDs, which are considered in the literature as a factor increasing the risk of bleeding [14]. As half of the patients received higher than recommended

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

doses of prophylaxis, we could logically expect a higher frequency of bleeding, but this was not observed. Indeed, inhibitors of the secretion of gastric acid (histamine H2 receptor and proton-pomp inhibitors) were given to 97.2% of patients, which could explain the lack of influence of NSAID or higher than recommended doses of prophylaxis administrations on major bleeding.

In the literature, computer-based systems designed to aid in clinical decision making have been promoted for their potential to improve the quality of health care by increasing physicians' performance [26–28]. Concerning the prevention of VTE, electronic alert systems have been shown to be effective in increasing physicians' use of prophylaxis and in reducing the incidence of VTE among hospitalized patients at risk [29]. After major orthopaedic procedures, the use of prophylaxis drugs is well-accepted by surgeons and, thus, the impact of electronic alerts systems on the use of prophylaxis is less important than for patients at moderate risk of VTE [27]. However, although physicians usually give VTE prophylaxis drugs after major orthopaedic interventions, our study shows a tendency to give higher than recommended doses. Quality of care could therefore be improved by a computer-based prescription system including doses of pharmacological treatments administered, patient characteristics justifying the choice of doses (e.g. patient weight) and the associated guidelines. Such a system may be useful not only for patients undergoing major orthopaedic procedures but also for other pathologies.

The hospital studied recently implemented an electronic prescription system in which the dose prescribed by the physician, the dose actually administered by the nurses and patient characteristics, such as patient weight, are detailed. More exhaustive and accurate pharmacological information than data from billing systems will be available in such a system. However, data input for the period analysed was not yet systematic, so it was not possible to use the data for the purposes of this study.

A weakness of our analysis is that we had no information on what happened in the 3 months following the in-hospital stays in terms of care delivered outside the hospital. In Belgium, no diagnostic information is collected about outpatient visits but a centralization of drugs delivered by public pharmacies is made. Easier access to these data, taking into account confidentiality aspects, would be of great use for analyses concerning quality of care. It would have allowed us to identify pharmacological treatment noted as potential risk factors, e.g. the use of anticoagulant or anti-platelet drugs during the week before the hospitalization or the use of NSAIDs before surgery. Treatment given after the hospital stay would also have been identified.

Our analysis is also influenced by the limits of the classification used to code diagnoses. For example, the ICD-9-CM classification grouped congenital deficiencies of factors I, II, V, VII, X, XII, XIII under only one code. Very specific pathologies, such as elevated levels of factors VII, IX or XI, noted as risk factors for VTE, could thus not be identified. The completeness of the diagnoses encoded in the Medical MBDS database is also an important factor in this type of study. Indeed, it is possible that some diagnoses are undercoded. However, by forgetting to code VTE or comorbidities, the severity level of the stay would be underestimated and consequently underfinanced. Hospitals thus have a financial interest to code these events.

To conclude, the standardized administrative databases could be used to detect some events, like symptomatic VTE after major orthopaedic surgery. However, evaluation of the prophylaxis given required access to medical records and internal billing data, mainly to identify patient weight and data on non-reimbursed drugs. This study demonstrates that the protocol for VTE prophylaxis was not always respected. This reinforces the need to develop informative tools that enable better monitoring of quality of care. Improved integration of in-hospital information systems may allow us to develop a form of 'alert' to detect patients at risk of VTE, not only when hospitalized for major orthopaedic surgery but for all other pathologies. At a national level, better integration of databases concerning outpatient drug consumption would be useful to complete quality assessment based on the medical MBDS database.

Acknowledgements

Isabelle Roch and Isabelle Pouplier (Researchers, Université catholique de Louvain, Public Health School, CIES-SESA, Brussels, Belgium) took part in the construction of the databases, Laurence Habimana (Biostatistics expert. Université catholique de Louvain, Public Health School, EPID, Brussels, Belgium) participated in the statistical analysis and Philippe Hainaut, PhD (Head of General Internal Medicine Department, Université catholique de Louvain, Cliniques universitaires Saint-Luc, Brussels, Belgium), contributed to the clinical aspect of the manuscript.

References

1. Healey, M. & Deverka, P. (1999) Panel 1: methodological issues in pharmacoeconomic evaluations – clinical studies. *Value in Health*, 2 (2), 73–77.
2. Arnold, R. G. & Kotsanos, J. G. (1999) Panel 3: methodological issues in conducting pharmacoeconomic evaluations – retrospective and claims database studies. *Value in Health*, 2 (2), 82–87.
3. Hay, J. (1999) Health care costs and outcomes: how should we evaluate real world data? *Value in Health*, 2 (6), 417–419.
4. Iezzoni, L. I. (1997) Assessing quality using administrative data. *Annals of International Medicine*, 127 (8 Part 2), 666–674.
5. Vincent, C., Davy, C., Esmail, A., Neale, G., Elstein, M., Cozens, J. F. & Walshe, K. (2006) Learning from litigation. The role of claims analysis in patient safety. *Journal of Evaluation in Clinical Practice*, 12 (6), 665–674.
6. Blumenthal, D. (1996) Part 1: quality of care – what is it? *New England Journal of Medicine*, 335 (12), 891–894.
7. Turpie, A. G., Bauer, K. A., Eriksson, B. I. & Lassen, M. R. (2002) Fondaparinux vs enoxaparin for the prevention of venous thromboembolism in major orthopedic surgery: a meta-analysis of 4 randomized double-blind studies. *Archives of International Medicine*, 162 (16), 1833–1840.
8. Annemans, L., Minjoulat-Rey, M. C., De Knock, M., Vranckx, K., Czarka, M., Gabriel, S. & Haentjens, P. (2004) Cost consequence analysis of fondaparinux versus enoxaparin in the prevention of venous thromboembolism after major orthopaedic surgery in Belgium. *Acta Clinica Belgica*, 59 (6), 346–357.
9. Geerts, W. H., Pineo, G. F., Heit, J. A., Bergqvist, D., Lassen, M. R., Colwell, C. W. & Ray, J. G. (2004) Prevention of venous thromboembolism: the Seventh ACCP Conference on Antithrombotic and Thrombolytic therapy. *Chest*, 126 (3 Suppl.), 338S–400S.
10. Colwell, C. W., Collis, D. K., Paulson, R., McCutchen, J. W., Bigler, G. T., Lutz, S. & Hardwick, M. E. (1999) Comparison of enoxaparin and warfarin for the prevention of venous thromboembolic disease after total hip arthroplasty. Evaluation during hospitalization and three months after discharge. *Journal of Bone and Joint Surgery. American Volume*, 81 (7), 932–940.

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

11. Haentjens, P. (1996) Thromboembolic prophylaxis in orthopaedic trauma patients: a comparison between a fixed dose and an individually adjusted dose of a low molecular weight heparin (nadroparin calcium). *Injury*, 27 (6), 385–390.
12. Tran, A. H. & Lee, G. (2003) Fondaparinux for prevention of venous thromboembolism in major orthopedic surgery. *Annals of Pharmacotherapy*, 37 (11), 1632–1643.
13. Kearon, C. (2003) Duration of venous thromboembolism prophylaxis after surgery. *Chest*, 124 (6 Suppl.), 386S–392S.
14. Comité d'évaluation des pratiques médicales en matière de médicaments. (2002) L'usage Adéquat de L'héparine Non Fractionnée, Des Héparines de Bas Poids Moléculaire et Des Anticoagulants Oraux Dans la Prévention et le Traitement de la Maladie Thrombo-Embolique Veineuse. Brussels: Institut national d'assurance maladie-invalidité.
15. Hainaut, P. (2003) Epidémiologie et pathogénie de la maladie thromboembolique veineuse. *Vaisseaux, Coeur, Poumons*, 8 (4), 147–151.
16. Thromboembolic Risk Factors (THRIFT) Consensus Group. (1992) Risk of and prophylaxis for venous thromboembolism in hospital patients. *BMJ*, 305 (6853), 567–574.
17. Caffo, B. S. & Booth, J. G. (2003) Monte carlo conditional inference for log-linear and logistic models: a survey of current methodology. *Statistical Methods in Medical Research*, 12 (2), 109–123.
18. Corcoran, C., Mehta, C., Patel, N. & Senchaudhuri, P. (2001) Computational tools for exact conditional logistic regression. *Statistics in Medicine*, 20 (17–18), 2723–2739.
19. Bauer, K. A., Eriksson, B. I., Lassen, M. R. & Turpie, A. G. (2001) Fondaparinux compared with enoxaparin for the prevention of venous thromboembolism after elective major knee surgery. *New England Journal of Medicine*, 345 (18), 1305–1310.
20. Eriksson, B. I., Bauer, K. A., Lassen, M. R. & Turpie, A. G. (2001) Fondaparinux compared with enoxaparin for the prevention of venous thromboembolism after hip-fracture surgery. *New England Journal of Medicine*, 345 (18), 1298–1304.
21. Lassen, M. R., Bauer, K. A., Eriksson, B. I. & Turpie, A. G. (2002) Postoperative fondaparinux versus preoperative enoxaparin for prevention of venous thromboembolism in elective hip-replacement surgery: a randomised double-blind comparison. *Lancet*, 359 (9319), 1715–1720.
22. Turpie, A. G., Bauer, K. A., Eriksson, B. I. & Lassen, M. R. (2002) Postoperative fondaparinux versus postoperative enoxaparin for prevention of venous thromboembolism after elective hip-replacement surgery: a randomised double-blind trial. *Lancet*, 359 (9319), 1721–1726.
23. Oster, G., Ollendorf, D. A., Vera-Llonch, M., Hagiwara, M., Berger, A. & Edelsberg, J. (2004) Economic consequences of venous thromboembolism following major orthopedic surgery. *Annals of Pharmacotherapy*, 38 (3), 377–382.
24. White, R. H., Romano, P. S., Zhou, H., Rodrigo, J. & Bargar, W. (1998) Incidence and time course of thromboembolic outcomes following total hip or knee arthroplasty. *Archives of International Medicine*, 158 (14), 1525–1531.
25. White, R. H., Zhou, H. & Romano, P. S. (2003) Incidence of symptomatic venous thromboembolism after different elective or urgent surgical procedures. *Thrombosis and Haemostasis*, 90 (3), 446–455.
26. Bates, D. W. & Gawande, A. A. (2003) Improving safety with information technology. *New England Journal of Medicine*, 348 (25), 2526–2534.
27. Durieux, P., Nizard, R., Ravaud, P., Mounier, N. & Lepage, E. (2000) A clinical decision support system for prevention of venous thromboembolism: effect on physician behavior. *JAMA: the Journal of the American Medical Association*, 283 (21), 2816–2821.
28. Hunt, D. L., Haynes, R. B., Hanna, S. E. & Smith, K. (1998) Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA: the Journal of the American Medical Association*, 280 (15), 1339–1346.
29. Kucher, N., Koo, S., Quiroz, R., Cooper, J. M., Paterno, M. D., Soukonnikov, B. & Goldhaber, S. Z. (2005) Electronic alerts to prevent venous thromboembolism among hospitalized patients. *New England Journal of Medicine*, 352 (10), 969–977.

© 2008 The Authors. Journal compilation © 2008 Blackwell Publishing Ltd

3 CONCLUSION AND PERSPECTIVES

The analysis of the quality of LMWH prophylaxis after major orthopedic surgery has demonstrated that the hospital protocol was not always respected. Half the patients received higher prophylaxis doses than advised and 11% received lower doses than advised. Such results implied a series of clinical practice recommendations.

In a first stage, the reasons of the low adherence of clinicians to the protocol should be discussed and then, a quality improvement cycle should be undergone. The aim of such a cycle is to continuously improve a process and is composed of four stages: to plan a change, to do the change, to check the change and finally to act (See figure 1) (1).

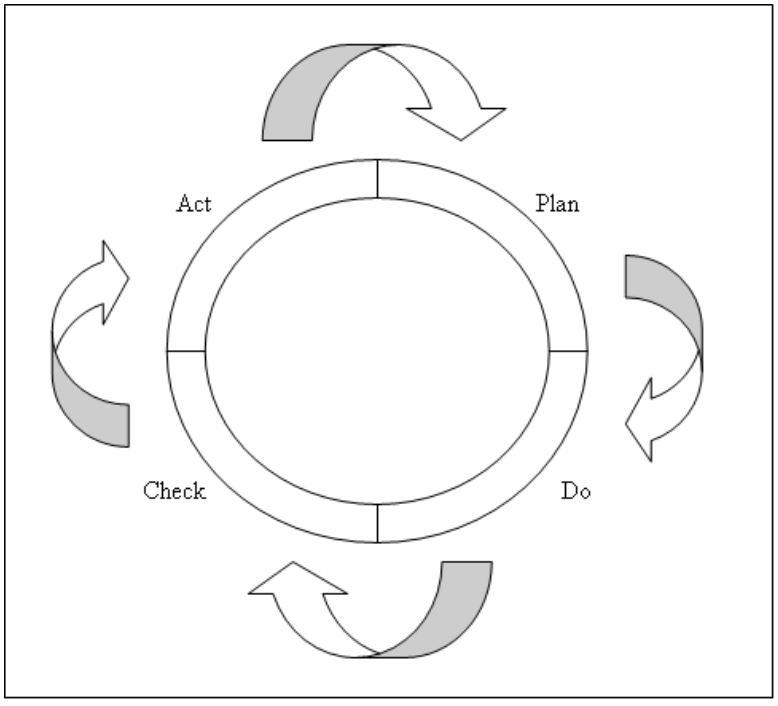

Figure 1 : Quality improvement cycle

In the context of our study, the following quality improvement cycle could be undergone. Firstly, an assessment of the hospital protocol in collaboration with the clinicians and according to international guidelines should be made and the reasons of the non adherence of clinicians should be highlighted. Then, in the **planning stage**, solutions to increase their adherence should be defined and included in a new protocol. In the **"do" stage**, the new protocol should be implemented and tested in the hospital. In the **"check stage"**, the adherences of clinicians and the incidences of VTE and side effects with this new protocol should be compared with results of the previous protocol. Finally, in the **"act" phase**, actions should be taken in concordance with the results of the study. So, if the adherence of the protocol and the incidence of events is not improved, others solutions should be taken and the cycle begins again. Otherwise, the protocol should be implemented and tested in a larger scale.

The need to develop computer-based systems that improve the monitoring of quality of care is also highlighted in our study. The creation of a kind of alert to detect patients at risk of VTE should be implemented, not only for patients undergoing major orthopedic surgery but also for other surgical and medical interventions where patients could be at risk.

On the other hand, even if the hospital's protocol was not respected, no significant impact on events was showed. A first explanation is the administration of inhibitors of gastric acid secretions to almost every patient (97.2%), limiting the risk of gastrointestinal bleeding. A second explanation may also be the fact that, even if some patients received superior doses compared to those advised by the protocol, it

remained within prophylactic ranges. The risk of bleeding was thus limited compared to patients that received curative doses (where doses are almost four times superior).

Because no significant impact on events was showed and because the impact in term of cost was minor, decision makers in hospitals have few incentives to implement such a quality improvement cycle. On the other hand, since July 2006, Belgium has set up a system of lump-sum payment for some drug specialties administrated to patients during the hospitalization, including LMWH prophylaxis drugs. With such a system, the interest for hospitals to implement quality improvement cycles could increase.

From a methodological point of view, the analysis of the databases' weaknesses has shown that improving could also be made. Indeed, in the studied databases, non-reimbursed pharmacological products were regrouped under a unique code. It was thus impossible to determine which kind of non-reimbursed products was administrated. On the other hand, such information was available in billing data of the hospitals under a national code, the CNK code. The possibility to systematically collect non-reimbursed products in these databases should thus be investigated.

Furthermore, it should be noted that some data having an impact on the risk of bleeding were not measured in our study. For instance, the administration of anti-depressive drugs, a bleeding risk factors, was not investigated. Moreover, if patients took hormonal treatments during their hospitalization, they usually brought their treatment from home and were not identified as patients at risk from the databases. Nevertheless, it should be noted that the mean age of patients in the sample was 70.3 years. Thus, the number of patients under hormonal

treatment not identified as patients at risk should be minor. Then, important data such as the patients' weight and pharmacological treatments delivered before and after the hospitalization were also not available in these hospital databases. An improved linkage between all other existing databases, with a unique patient identification, and especially a linkage with data from the patients' medical electronic records and data on outpatient drug consumption would considerably increase the quality of economic evaluations performed through Belgian databases. Indeed, lack of data is a major drawback for studies performed retrospectively through databases but with an improved linkage, this drawback would be limited. As a model, the Swedish health care system has over 50 clinical databases called "quality registries" to assess health care quality (2). Nevertheless, it should be noted that it is important to keep the identification of patient confidential in order to preserve the privacy of his medical information.

On the other hand, the current Belgian hospital databases included data on reimbursed pharmacological products administrated to the patients. This information allowed us to determine which prophylaxis was given to the patients, which was not the case in other studies assessing the incidence of VTE events from administrative databases (3-5). These databases also permitted to identify the incidence of clinical VTE events, major bleeding and death from any causes during the hospitalization but also during readmission in the same hospitals. Direct health care costs, in term of fees but not as real costs, could also be measured. Thus, these data would allow one to perform a full economic evaluation. However, during the period studied, only one LMWH prophylaxis was identified, i.e. nadroparin.

Enoxaparin was only introduced in the hospital the following years. A comparison of these two prophylaxis drugs in term of incremental cost-effectiveness ratio would thus be interesting to make in a future research.

Moreover, a comparison of incidences between hospitals in order to establish benchmarks could also be interesting to perform in the future from these hospital databases. Such a process would allow one to manage the quality of clinical practice by measuring the efficiency of hospitals, by comparing in details the practices in “efficient” and “inefficient” hospitals and by setting up a continuous quality improvement process. Nevertheless, it should be noted that such an analysis would allow us to detect abnormalities but that these abnormalities would have to be analyzed in detail before to make any conclusion. Indeed, such retrospective analyses would be based on administrative databases and thus may not identify some confounding factors.

References

(1) Andreoni V, Bilak Y, Bukumira M, Halfer D, Lynch-Stapleton P, Perez C. Project management: putting continuous quality improvement theory into practice. J Nurs Care Qual 1995 Apr;9(3):29-37.

(2) Sousa P, Bazeley M, Johansson S, Wijk H. The use of national registries data in three European countries in order to improve health care quality. Int J Health Care Qual Assur Inc Leadersh Health Serv 2006;19(6-7):551-60.

(3) Oster G, Ollendorf DA, Vera-Llonch M, Hagiwara M, Berger A, Edelsberg J. Economic consequences of venous thromboembolism following major orthopedic surgery. Ann Pharmacother 2004 Mar;38(3):377-82.

(4) White RH, Romano PS, Zhou H, Rodrigo J, Bargar W. Incidence and time course of thromboembolic outcomes following total hip or knee arthroplasty. Arch Intern Med 1998 Jul 27;158(14):1525-31.

(5) White RH, Zhou H, Romano PS. Incidence of symptomatic venous thromboembolism after different elective or urgent surgical procedures. Thromb Haemost 2003 Sep;90(3):446-55.

PART 2 : THE USE OF MARKOV MODEL

1 METHODOLOGICAL CHOICES AND RESEARCH QUESTION

Because Markov models are especially adapted to assess the treatment of chronic diseases, we decided to investigate the treatment of patients with chronic hepatitis C. In Belgium, guidelines advise to treat patients with moderate chronic hepatitis C and elevated alanine transaminase (ALT) levels (1). However, some questions remain, firstly concerning the treatment of patients with mild chronic hepatitis C and elevated ALT, and secondly concerning the treatment of patients with moderate chronic hepatitis C and persistently normal alanine transaminase (PNALT) levels. So, in order to respond to these questions and to assess their impact, two Markov models were created.

For patients with mild chronic hepatitis C, it is currently advised to only follow these patients and to treat them only once the stage of moderate chronic hepatitis C was reached. However, the question of a direct treatment of these patients was raised and the impact of such strategy on long term costs and effects is unknown (2).

Thus, this first model compared two strategies:

- A follow-up of patients with mild chronic hepatitis C and elevated ALT and a treatment once the stage of moderate chronic hepatitis C was reached.
- A direct treatment of patients with mild chronic hepatitis C and elevated ALT.

For patients with moderate chronic hepatitis C and PNALT, it is currently advised to not treat them but the question of treating them is also under discussion (3). In the second model, the strategies to treat or not to treat patients with moderate chronic hepatitis C and PNALT were thus compared.

Moreover, because clinical data on the treatment of chronic hepatitis C were provided by RCTs and not by observational studies, we also tested under which conditions the treatment would become inefficient in daily practice, i.e. with “effectiveness” data instead of “efficacy” data.

Concerning the models’ parameters, different choices were made. The first difference concerned the utilities. In the first model, higher utility values were used. Even if these values came from a study where the estimation method was not clear (4), we decided to use them. Such a choice was based on the conservative approach that in the first model, utilities had an important impact on results and that higher utilities were in favor of no direct treatment.

Annual transitions probabilities between fibrosis scores also differed in the two models. In the second model, a distinction of transition

probabilities according to the gender of the patients was made, even if model stratification by gender had no impact on the final result.

A list of all abbreviations used in this part can be found in table 1.

Table 1 : List of abbreviations

ALT	Alanine transaminase
EVR	Early viral response
FPR	Fibrosis progression rate
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
ICER	Incremental cost-effectiveness ratio
PCR	Polymerase chain reaction
PEG IFN	Pegylated interferon
PNALT	Persistently normal alanine transaminase
QALY	Quality-adjusted life-year
RIZIV/INAMI	National health institute (Rijksinstituut voor ziekte- en invaliditeitsverzekering / Institut national d'assurance maladie-invalidité)
RNA	Ribonucleic acid
SVR	Sustained viral response
VAT	Value added tax

References

(1) Michielsen P, Brenard R, Bourgeois N, de GC, Delwaide J, Henrion J, et al. Hepatitis C: screening, treatment and prevention practical guidelines. Acta Gastroenterol Belg 2003 Jan;66(1):15-9.

(2) Levine RA. Treating histologically mild chronic hepatitis C: monotherapy, combination therapy, or tincture of time? Ann Intern Med 1998 Aug 15;129(4):323-6.

(3) Shiffman ML, Diago M, Tran A, Pockros P, Reindollar R, Prati D, et al. Chronic hepatitis C in patients with persistently normal alanine transaminase levels. Clin Gastroenterol Hepatol 2006 May;4(5):645-52.

(4) Siebert U, Sroczynski G, Rossol S, Wasem J, Ravens-Sieberer U, Kurth BM, et al. Cost effectiveness of peginterferon alpha-2b plus ribavirin versus interferon alpha-2b plus ribavirin for initial treatment of chronic hepatitis C. Gut 2003 Mar;52(3):425-32.

2 A HEALTH ECONOMIC MODEL TO ASSESS THE COST-EFFECTIVENESS OF PEG IFN α-2A AND RIBAVIRIN IN PATIENTS WITH MILD CHRONIC HEPATITIS C

S Gerkens, M Nechelput, L Annemans, B Peraux, M Mouchart, C Beguin, Y Horsmans. A health economic model to assess the cost-effectiveness of PEG IFN α-2a and ribavirin in patients with mild chronic hepatitis C. Journal of viral hepatitis 2007 Aug;14(8):523-36.

NB: The definitive version is available at www.blackwell-synergy.com.
© 2007 Blackwell Publishing Ltd.

Journal of Viral Hepatitis, 2007, **14**, 523–536 doi:10.1111/j.1365-2893.2006.00831.x

A health economic model to assess the cost-effectiveness of PEG IFN α-2a and ribavirin in patients with mild chronic hepatitis C

S. Gerkens,[1] M. Nechelput,[2] L. Annemans,[3] B. Peraux,[2] M. Mouchart,[4] C. Beguin[1] and Y. Horsmans[1] [1]*Université catholique de Louvain, Cliniques universitaires Saint-Luc, Brussels;* [2]*NV ROCHE SA Brussels;* [3]*Universiteit Gent, Vrije Universiteit Brussel, Brussels; and* [4]*Université catholique de Louvain, Louvain-La-Neuve, Belgium*

Received June 2006; accepted for publication October 2006

SUMMARY. According to the current guidelines, it is advised not to treat patients with mild chronic hepatitis C. However, discussions as to giving immediately a treatment (direct treatment) to these patients have started and the incremental cost-effectiveness ratio (ICER) of such strategy is still unknown. The aim of this study was to estimate, in the health care payer perspective, the ICER of a direct treatment of patients with mild chronic hepatitis C in comparison with the strategy of monitoring these patients and treat them when the disease will progress to the state of moderate chronic hepatitis. The treatment assessed was the current standard treatment composed of pegylated interferon α-2a and ribavirin. At the beginning of the study, patients were aged 45. Long-term economic and clinical outcomes over a 30-year period were predicted using a Markov simulation model. Data were obtained from published literature. Monte Carlo simulations were used to determine 95% confidence intervals of results. The ICER of a direct treatment with PEG IFN α-2a and ribavirin is €23 046/QALY (CI 95% €3 882-€42 392) for genotypes 1–4–5–6 and €4 631/QALY (CI 95% €797-€7 881) for genotypes 2–3. Sensitivity analysis shows that it is only in extreme circumstances related to the utilities that the ICER for genotypes 1–4–5–6 is unacceptably high for the society (>€50 000). Even though a direct treatment is more expensive, it gives the advantage of curing greater number of patients and of increasing quality-adjusted life-years (QALYs), implying that such strategy is generally cost-effective at a threshold of €50 000/QALY.

Keywords: cost-effectiveness, economic, Markov model, mild chronic hepatitis C, pegylated interferon.

INTRODUCTION

Hepatitis C virus (HCV) is a major public health problem. Worldwide, it is estimated that 170 million persons are infected, i.e. 3% of the world population [1]. In Belgium, the prevalence is estimated at nearly 1% [2]. People infected by HCV can remain asymptomatic during more than 20 years and 35–80% of infected people are unaware of their seropositivity. Moreover, when the infection is diagnosed, few patients are treated [3]. However, 80% of the infected patients will develop chronic hepatitis C, which will evolve for 20% of them to cirrhosis 20–30 years later. Afterwards, 3.9%of patients per year will develop a decompensated cirrhosis and 1.4% per year a hepatocellular carcinoma [4,5]. In Belgium, more than 29% of the indications for liver transplantations are related to hepatitis C [6].

The severity of chronic hepatitis C can be classified into three categories, according to the type and the amount of fibrosis of the liver. Mild chronic hepatitis C is characterized by no or little fibrosis, located in the portal area and without septa (METAVIR fibrosis scores of F0-F1). Moderate chronic hepatitis C shows portal or periportal fibrosis with some septa (METAVIR fibrosis score of F2) and severe chronic hepatitis C is determined by the presence of septal fibrosis without cirrhosis (METAVIR fibrosis score of F3) [7,8].

Presently, no vaccine is available but curing the disease is possible. The current standard therapy for patients with moderate to severe hepatitis C is the combination of pegylated interferon with ribavirin [1,9].

The best therapeutic option for patients with mild chronic hepatitis C is still debated [10,11]. Indeed, these patients will not always develop a complication and the cost of treatment for all these patients is expected to be high. The current advice is the follow-up of these patients and to initiate treatment only when the state of moderate to severe chronic

Abbreviations: EVR, early viral response; SVR, sustained viral response; PCR, polymerase chain reaction; HCV, hepatitis C virus; ICER, incremental cost-effectiveness ratio; IFN, interferon; QALYs, quality-adjusted life-years.

Correspondence: Horsmans Yves, Gastroenterology Department, Cliniques universitaires Saint-Luc, Université catholique de Louvain, 10 Avenue Hippocrate, Box 2822, B 1200 Brussels, Belgium. E-mail: sophie.gerkens@infm.ucl.ac.be

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

hepatitis C is reached [9]. However, taking into account the long-term complications, the effectiveness and the cost of treatment of patients with a mild disease would be interesting.

The goal of this study was to compare two strategies for patients with mild chronic hepatitis C and Metavir fibrosis score F1: direct treatment or absence of treatment until the disease evolves to METAVIR F2 fibrosis score.

With the growing cost of health care, efficacy and safety of drugs become insufficient for taking a good decision. Today, the priority is to reduce health costs without deteriorating the quality of care, or to improve quality of care at a reasonable cost [12]. Consequently, the effectiveness but also the costs of these two strategies were investigated in this study.

METHODS AND DESIGN

The Markov model

The incremental cost-effectiveness ratio (ICER) of these two strategies was investigated using a Markov simulation model representing the natural history of chronic hepatitis C. This technique allows modelling the evolution of a cohort of patients over time and assumes that individuals are always in one of a finite number of health states, called 'Markov state'. They progress in a fixed unit of time, referred to as 'Markov cycle'. For each Markov state, a fixed annual transition probability towards the following state, a cost and a utility are referred [13,14]. Annemans et al. recently developed a model for patients with chronic hepatitis C based on the literature [6]. Within the framework of this study, this model was used and adapted in order to distinguish between the various states of chronic hepatitis C, i.e. the scores of fibrosis.

In this study, annual cycles were considered, meaning that evolution or not to the following state of the disease will be evaluated on a yearly basis. Because in real-life patients do not progress between states at the beginning or the end of a cycle, a half-cycle correction was performed.

At the beginning of the study, all patients had mild chronic hepatitis C (METAVIR fibrosis score of F1) and were on average 45 years old. Younger and older patients were also tested and the impact of age on the result was included in the discussion.

Two therapeutic strategies were assessed by this model: to directly treat the patients (first strategy) or to monitor the patients and to treat them only when they reached the state of moderate chronic hepatitis C (second strategy). The programming was made in Microsoft excel.

For genotypes 1–4–5–6, the treatment assessed was PEG IFN α-2a 180 μg once a week in combination with ribavirin 1000–1200 mg/day for 48 weeks. After 12 weeks of treatment, an early predictive test was performed. If the viral load was not reduced by more than 2 $\log_{10}$, the treatment was stopped. In the analysis, the duration of treatment considered in these cases was 16 weeks and not 12, in order to take into account the time necessary to obtain the results and to make the decision to stop the therapy. This early viral response (EVR) was determined by a quantitative test (Cobas Amplicor HCV Monitor test; Roche Diagnostics, Mannheim, Germany) which measured the reduction of HCV RNA and

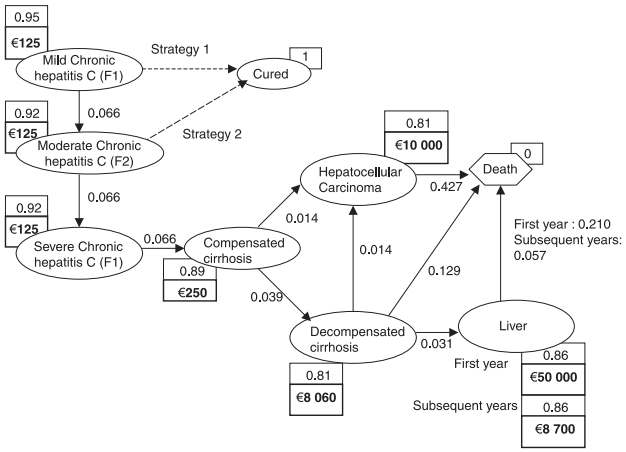

Fig. 1 Markov model for HCV.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

was positive when the test showed a reduction of more than 2 $\log_{10}$ or unquantifiable HCV RNA.

For genotypes 2 and 3, the treatment considered was PEG IFN α-2a 180 μg once a week in combination with ribavirin 800 mg/day for 24 weeks. For these genotypes, no early predictive test was performed.

For each genotype, the sustained viral response (SVR) rate was calculated 24 weeks post-treatment cessation and was defined as a polymerase chain reaction (PCR) negative test (absence of HCV RNA; Cobas Amplicor HCV test). Patients with SVR are regarded as cured from the disease. In the absence of SVR, the patients could evolve to compensated cirrhosis and thereafter to decompensated cirrhosis or hepatocellular carcinoma. Patients with decompensated cirrhosis could die, develop a hepatocellular carcinoma or undergo a liver transplantation. Finally, patients with hepatocellular carcinoma or liver transplantation could also die (Fig. 1). As in the model developed by Annemans et al. [6], it was assumed that patients with compensated cirrhosis could not directly progress towards death and that a patient could not be found in two different states within the same year.

The viral response rates (SVR and EVR)

The SVR rates used in our model came from the study completed by Annemans et al. [6]. For genotypes 1–4–5–6, they were based on the study published by Fried et al. [15]. This randomized clinical trial compared two types of treatments: PEG IFN α-2a 180 μg once a week in combination with ribavirin 1000–1200 mg/day for 48 weeks and interferon α-2b 3 MIU three times per week in combination with ribavirin 1000–1200 mg/day for 48 weeks. Among the 311 patients with genotypes 1–4–5–6 treated, 253 patients had an EVR at 12 weeks and on these patients, 138 had an SVR 24 weeks after the end of the treatment, i.e. an EVR rate of 81.4% and an SVR rate of 44.4% (Table 1).

Table 1 EVR and SVR rates for genotypes 1–4–5–6 [15]

	SVR	No SVR	Total	EVR%
EVR 12 weeks	138	115	253	81.4%
No EVR 12 weeks	1	57	58	
Total	139	172	311	
SVR% (if EVR)	44.4%			

EVR, early viral response; SVR, sustained viral response.
Genotypes 1–4–5–6: peginterferon alfa-2a 180 μg/week and ribavirin 1000–1200 mg/day for 48 weeks.

For patients with genotypes 2 and 3 and treated with PEG IFN α-2a 180 μg once a week in combination with ribavirin 800 mg/day for 24 weeks, Annemans et al. [6] calculated the SVR rate from two different studies, the study of Hadziyannis et al. [16] and the study of Fried et al. [15]. They used a hazard rate technique and obtained an SVR rate of 80.1% [6]. It was supposed that the SVR rate for patients with mild chronic hepatitis C was the same as for those with a moderated chronic hepatitis C, which was a conservative approach.

Annual transition probabilities

Alberti [7] suggested that, after 30 years, 60% of people with a fibrosis score of F1 and elevated ALT will develop severe chronic hepatitis C (fibrosis score of F3), or even cirrhosis (fibrosis score of F4), natural deaths excluded [7]. This result was used to estimate the annual transition probabilities between the fibrosis scores of F1-F2, F2-F3, F3-F4 of the model. As in the study realized by Salomon et al.

Table 2 Annual transition probabilities

From	To	%	References
Mild chronic hepatitis C (METAVIR fibrosis score F1)	Moderate chronic hepatitis C (METAVIR fibrosis score F2)	6.6 (2.1–8.2)	Alberti [7], Bennett et al. [19]
Moderate chronic hepatitis C (METAVIR fibrosis score F2)	Severe chronic hepatitis C (METAVIR fibrosis score F3)	6.6 (2.1–8.2)	Alberti [7], Bennett et al. [19]
Severe chronic hepatitis C (METAVIR fibrosis score F3)	Compensated cirrhosis	6.6 (1.0–23.2)	Younossi et al. [18]
Compensated cirrhosis	Decompensated cirrhosis	3.9 (2.0–8.3)	Younossi et al. [18]
	Hepatocellular carcinoma	1.4 (1.0–4.4)	Younossi et al. [18]
Decompensated cirrhosis	Hepatocellular carcinoma	1.4 (1.0–4.4)	Younossi et al. [18]
	Liver transplantation	3.1 (1.0–6.2)	Younossi et al. [18]
	Death	12.9 (6.5–19.3)	Younossi et al. [18]
Hepatocellular carcinoma	Death	42.7 (33–86)	Younossi et al. [18]
Liver transplantation	Death (first year)	21.0 (6–42)	Younossi et al. [18]
	Death (subsequent years)	5.7 (2.4–11)	Younossi et al. [18]

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Table 3 Utility estimates

Markov state	Younossi et al. [18]	Bennett et al. [19]	Sagmeister et al. [20]	Sennfalt et al. [21]	Siebert et al. [22]	Model
SVR	1	1		1		1
Utility multiplier for peg-interferon + ribavirin					0.90 (0.84–0.96)	0.90 (0.84–0.96)
Chronic hepatitis C	0.82 (0.6–0.9)		0.90			
Mild		0.82 (0.63–0.98)		0.82	0.95 (0.90–1.00)	0.95 (0.63–1.00)
Moderate to severe		0.78 (0.6–0.9)		0.78	0.92 (0.89–0.95)	0.92 (0.60–0.95)
Compensated cirrhosis	0.78 (0.5–0.9)	0.7 (0.5–0.9)	0.75	0.70	0.89 (0.82–0.92)	0.89 (0.50–0.92)
Decompensated cirrhosis	0.65 (0.3–0.88)				0.81 (0.76–0.87)	0.81 (0.30–0.88)
Hepatocellular carcinoma	0.25 (0.1–0.5)	0.1 (0.2–0.5)	0.20	0.10	0.81 (0.76–0.87)	0.81 (0.10–0.87)
Liver transplantation, first year	0.5 (0.11–0.70)	0.5 (0.11–0.70)	0.70	0.50	0.86 (0.73–0.99)	0.86 (0.11–0.99)
Liver transplantation, subsequent years	0.7 (0.24–0.87)	0.7 (0.24–0.87)	0.80	0.70	0.86 (0.73–0.99)	0.86 (0.24–0.99)

Table 4 Cost estimates

	Costs per year					
Markov states	Wong et al. [24]	Sennfalt et al. [21]	Jusot and Colin [25]	Younossi 1999 [18]	Shepherd et al. [1]	Wong and Nevens [26]
Currency	$	$	€	$	£	€
Chronic hepatitis C			579	319	95	125
Compensated cirrhosis	110	163	1 400	425	237	250
Decompensated cirrhosis		6 885	11 500	22 100		8 060
Hepatocellular carcinoma			10 400		13 320	10 000
Liver transplantation						
First year	108 659	97 281	92 700	269 072		50 000
Subsequent years	18 976	16 093	7 900	25 910		8 700

Table 5 Baseline results for genotypes 1–4–5–6

Genotypes 1–4–5–6	First strategy	Second strategy	Difference
Cost			
Drug costs	€15931.58	€9648.14	€6283.44
Medical care costs	€2113.92	€2735.50	€−621.58
Total costs	€18045.50	€12383.64	€5661.86
Efficacy			
Discounted quality-adjusted life years (QALY)	20.56	20.32	0.25
% cured people after 30 years	44.40	35.57	8.83
Cost-efficacy rate			
Cost/QALY	€877.64	€609.56	€268.08
Incremental cost-efficacy ratio			€23046.07/QALY

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

[17], these fibrosis scores were supposed to be constant. The other probabilities came from the model developed by Annemans et al. [6], themselves coming from a study of Younossi et al. [18] (Table 2). The mortality due to natural causes was also included in the model using age-specific mortality rates [6].

Table 6 Baseline results for genotypes 2–3

Genotypes 2–3	First strategy	Second strategy	Difference
Cost			
Drug costs	€8427.08	€5103.43	€3323.65
Medical care costs	€756.60	€1877.96	€−1121.36
Total costs	€9183.68	€6981.39	€2202.29
Efficacy			
Discounted quality-adjusted life years (QALY)	21.20	20.73	0.48
% cured people after 30 years	80.10	64.18	15.92
Cost-efficacy rate			
Cost/QALY	€433.16	€336.85	€96.31
Incremental cost-efficacy ratio			€4630.56/QALY

Table 7 Sensitivity analysis on annual transition probabilities

Genotypes 1–4–5–6	Minimal value	Baseline	Maximal value
First strategy			
Cost			
Drug costs	€15931.58	€15931.58	€15931.58
Medical care costs	€1257.39	€2113.92	€4089.03
Total costs	€17188.98	€18045.50	€20020.62
Efficacy			
Discounted quality-adjusted life years (QALY)	20.81	20.56	19.80
Second strategy			
Cost			
Drug costs	€4625.32	€9648.14	€10594.59
Medical care costs	€2053.80	€2735.50	€4769.43
Total costs	€6679.12	€12383.64	€15364.02
Efficacy			
Discounted QALY	20.48	20.32	19.51
Incremental cost-efficacy			
Cost/QALY	€31978.49	€23046.07	€16000.71
Genotypes 2–3			
First strategy			
Cost			
Drug costs	€8427.08	€8427.08	€8427.08
Medical care costs	€450.04	€756.60	€1,463.52
Total costs	€8877.12	€9183.68	€9890.60
Efficacy			
Discounted QALY	21.29	21.20	20.93
Second strategy			
Cost			
Drug costs	€2446.58	€5103.43	€5604.05
Medical care costs	€1886.81	€1877.96	€2690.98
Total costs	€4333.39	€6981.39	€8295.04
Efficacy			
Discounted QALY	20.63	20.73	20.38
Incremental cost-efficacy			
Cost/QALY	€6871.48	€4630.56	€2897.84

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Utilities

The utilities represent the preferences of individuals for a specific health state. They vary from 0, which represents death, to 1, which represents the perfect health state. These utilities were used to estimate the health-related quality of life and then the ICER in terms of cost per QALY. For both strategy investigated, a QALY was calculated by multiplying the time spent in each health state by the utility of the state and then by summing the whole over the life expectancy [12]. In the literature, several estimates of utilities could be found concerning the different states of the model [18–22]. In this study, utilities provided by Siebert et al. [22] were chosen as they were more elevated than those published in other studies [18–21], using a conservative approach (Table 3). Deterioration of the quality of life due to the treatment was included in the model using the utility multiplier for peg-interferon and ribavirin assessed by Siebert et al. [22]. These utilities were estimated from a transformed visual analogue scale, EuroQoL and doctor-based estimates.

In order to take into account the differences between the studies, a sensitivity analysis, which assessed the range of values coming from the literature, was performed. Because patients did not progress at the same time in the two strategies, the future utilities were discounted.

Complication costs

Various estimates of the costs were found in the literature [1,18,21,23–26]. Once more, the costs estimates used in this study came from the model of Annemans et al. [6] because they were derived from a Belgian study performed by Wong in 2002 [26] (Table 4). These costs, calculated according to current prices' in Belgium, were lower than those in studies

Table 8 Sensitivity analysis on utilities

Genotypes 1–4–5–6	Minimal value	Baseline	Maximal value
First strategy			
Cost			
Drug costs	€15931.58	€15931.58	€15931.58
Medical care costs	€2113.92	€2113.92	€2113.92
Total costs	€18045.50	€18045.50	€18045.50
Efficacy			
Discounted quality-adjusted life years (QALY)	16.70	20.56	21.04
Second strategy			
Cost			
Drug costs	€9648.14	€9648.14	€9648.14
Medical care costs	€2735.50	€2735.50	€2735.50
Total costs	€12383.64	€12383.64	€12383.64
Efficacy			
Discounted QALY	14.89	20.32	21.03
Incremental cost-efficacy			
Cost/QALY	€3130.52	€23046.07	€838217.33
Genotypes 2–3			
First strategy			
Cost			
Drug costs	€8427.08	€8427.08	€8427.08
Medical care costs	€756.60	€756.60	€756.60
Total costs	€9183.68	€9183.68	€9183.68
Efficacy			
Discounted QALY	19.82	21.20	21.37
Second strategy			
Cost			
Drug costs	€5103.43	€5103.43	€5103.43
Medical care costs	€1877.96	€1877.96	€1877.96
Total costs	€6981.39	€6981.39	€6981.39
Efficacy			
Discounted QALY	16.52	20.73	21.33
Incremental cost-efficacy			
Cost/QALY	€668.07	€4630.56	€48102.22

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

from other countries. Higher costs being beneficial for the cost-effectiveness of a direct treatment, the approach adopted was conservative.

The medical care costs of mild, moderate and severe chronic hepatitis were supposed to be the same, these patients requiring no different caring. Here also, costs were discounted.

Drug costs

The Belgian health care payer perspective was adopted in this study. So, the current public prices published by the INAMI (Institut national d'assurance maladie invalidité) in January 2006, VAT included and patient contribution deduced, were used. For genotypes 1–4–5–6, the treatment cost €18 187/patient. By taking into account the EVR rate, and thus the proportion of patient for which the treatment was effectively stopped after 16 weeks, the mean treatment cost was reduced to €15 932/patient. For genotypes 2–3, the treatment cost was equal to €8,427/patient.

In the first strategy, all patients were treated directly. With the second strategy, they were treated only when they progressed towards the state of moderate chronic hepatitis C. In order to take into account this temporal difference, the drugs costs were discounted.

Temporal characteristics of the study

The duration of the study of Annemans et al. was 25 years [6]. Given that the model developed in our analysis began with patients with mild chronic hepatitis C, the evolution of the disease was slower and the period taken into account in this study must be longer. Consequently, the duration selected for this model was 30 years. In order to include the temporal differences between the two strategies, future costs

Table 9 Sensitivity analysis on costs

Genotypes 1–4–5–6	Minimal value	Baseline	Maximal value
First strategy			
Cost			
Drug costs	€15931.58	€15931.58	€15931.58
Medical care costs	€1691.13	€2113.92	€2536.70
Total costs	€17622.72	€18045.50	€18468.28
Efficacy			
Discounted quality-adjusted life years (QALY)	20.56	20.56	20.56
Second strategy			
Cost			
Drug costs	€9648.14	€9648.14	€9648.14
Medical care costs	€2188.40	€2735.50	€3282.59
Total costs	€11836.54	€12383.64	€12930.74
Efficacy			
Discounted QALY	20.32	20.32	20.32
Incremental cost-efficacy			
Cost/QALY	€23552.08	€23046.07	€22540.05
Genotypes 2–3			
First strategy			
Cost			
Drug costs	€8427.08	€8427.08	€8427.08
Medical care costs	€605.28	€756.60	€907.92
Total costs	€9032.36	€9183.68	€9335.00
Efficacy			
Discounted QALY	21.20	21.20	21.20
Second strategy			
Cost			
Drug costs	€5103.43	€5103.43	€5103.43
Medical care costs	€1502.37	€1877.96	€2253.55
Total costs	€6605.80	€6981.39	€7356.98
Efficacy			
Discounted QALY	20.73	20.73	20.73
Incremental cost-efficacy			
Cost/QALY	€5102.12	€4630.56	€4159.00

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

were discounted at a rate of 3% and future utilities were discounted at a rate of 1.5%, as recommended by recent Belgian guidelines for cost-effectiveness analysis [27].

RESULTS

For genotypes 1–4–5–6

The Markov technique permitted to model the evolution of a cohort of patients with mild chronic hepatitis C over a 30-year period. At each state of the model, a cost and a utility were associated and thus, the total cost and the QALY of the strategies could be calculated.

After 30 years, 19% patients do not evolve towards the state of moderate chronic hepatitis C with the second strategy and are thus not treated. Moreover, with this second strategy, treated people do not receive the therapy at the beginning of the study and its cost is thus discounted. For these reasons, the drug costs €6 283 more with a direct treatment of patients with mild chronic hepatitis C.

On the other hand, the model shows that, with a direct treatment, medical care costs other than the therapy cost are reduced by €622 because more patients are cured and this more quickly. In terms of quality-adjusted life expectancy, people gain 0.25 QALY thanks to a direct treatment. Moreover, with this first strategy, there are additional 8.8% of patients cured.

Thus, even if the therapy is more expensive when patients are treated directly, medical care costs other than the therapy cost are reduced, more people are cured and the number of healthy years is higher, resulting in an ICER of €23 046/QALY (Table 5).

Table 10 Sensitivity analysis on SVR and EVR

Genotypes 1–4–5–6	Minimal value	Baseline	Maximal value
First strategy			
Cost			
Drug costs	€13957.72	€15931.58	€17905.45
Medical care costs	€2451.54	€2113.92	€1776.30
Total costs	€16409.25	€18045.50	€19681.75
Efficacy			
Discounted quality-adjusted life years (QALY)	20.41	20.56	20.71
Second strategy			
Cost			
Drug costs	€8452.77	€9648.14	€10843.52
Medical care costs	€2948.80	€2735.50	€2522.19
Total costs	€11401.57	€12383.64	€13365.71
Efficacy			
Discounted QALY	20.22	20.32	20.41
Incremental cost-efficacy			
Cost/QALY	€26089.95	€23046.07	€21094.78
Genotypes 2–3			
First strategy			
Cost			
Drug costs	€8427.08	€8427.08	€8427.08
Medical care costs	€1365.68	€756.60	€147.52
Total costs	€9792.76	€9183.68	€8574.60
Efficacy			
Discounted QALY	20.93	21.20	21.47
Second strategy			
Cost			
Drug costs	€5103.43	€5103.43	€5103.43
Medical care costs	€2262.77	€1877.96	€1493.15
Total costs	€7366.20	€6981.39	€6596.58
Efficacy			
Discounted QALY	20.55	20.73	20.90
Incremental cost-efficacy			
Cost/QALY	€6414.02	€4630.56	€3452.78

SVR, sustained viral response; EVR, early viral response.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

For genotypes 2–3

These patients are treated less longer and therefore the impact of the treatment cost is less important than with the other genotypes. Medical care costs other than the therapy cost are 2.5 times lower with a direct treatment (€757 *vs* €1 878), patients gain 0.48 QALY and this strategy permits to cure 16% more patients. Consequently, the ICER is better than for genotypes 1–4–5–6 and is €4 631/QALY (Table 6).

Sensitivity analysis

Considering the variations of input estimates from different sources, a sensitivity analysis was necessary. The key parameters of the model were the annual transition probabilities, the utilities, the costs, the SVR rates, the duration of the study, the age at the beginning of the study and the discount rate.

Sensitivity analysis on transition probabilities

The range of possible values of annual transition probabilities, determined by Younossi et al. [18] and by Bennett et al. [19], was assessed.

With a faster progression of the disease, the incremental cost/QALY rate improve. In all the situations considered, the medical care costs other than the therapy cost are lower with the direct treatment strategy and the quality-adjusted life expectancy is better. The ICERs range between €16 001 and €31 978 for genotypes 1–4–5–6 and between €2 898 and €6 871 for genotypes 2–3 (Table 7).

Sensitivity analysis on utilities

The range of possible utilities extracted from the literature [18–22] was used for this sensitivity analysis (see Table 3). The more the patients consider their state as 'unpleasant', the

Table 11 Sensitivity analysis on time frame

Genotypes 1–4–5–6	Minimal value	Baseline	Maximal value
First strategy			
Cost			
Drug costs	€15931.58	€15931.58	€15931.58
Medical care costs	€1258.41	€2113.92	€2803.27
Total costs	€17190.00	€18045.50	€18734.85
Efficacy			
Discounted quality-adjusted life years (QALY)	15.63	20.56	23.29
Second strategy			
Cost			
Drug costs	€8865.00	€9648.14	€9850.42
Medical care costs	€1796.07	€2735.50	€3455.45
Total costs	€10661.07	€12383.64	€13305.88
Efficacy			
Discounted QALY	15.42	20.32	23.02
Incremental cost-efficacy			
Cost/QALY	€32161.18	€23046.07	€20343.01
Genotypes 2–3			
First strategy			
Cost			
Drug costs	€8427.08	€8427.08	€8427.08
Medical care costs	€450.40	€756.60	€1,003.33
Total costs	€8877.48	€9183.68	€9430.41
Efficacy			
Discounted QALY	16.05	21.20	24.14
Second strategy			
Cost			
Drug costs	€4689.18	€5103.43	€5210.42
Medical care costs	€1420.37	€1877.96	€2179.90
Total costs	€6109.55	€6981.39	€7390.32
Efficacy			
Discounted QALY	15.64	20.73	23.62
Incremental cost-efficacy			
Cost/QALY	€6805.69	€4630.56	€3987.79

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

more the ICERs improve. On the other hand, if patients consider their state as little or no disturbing, a direct treatment become not any more cost-effective. However, irrespective of the value of utility assessed, the quality-adjusted life expectancy remains higher with a direct treatment, even if the difference is reduced with an increase of utility values. The ICERs range between €3 131 and €838,217 for genotypes 1–4–5–6 and between €668 and €48 102 for genotypes 2–3 (Table 8).

Sensitivity analysis on costs

For the sensitivity analysis on the costs, a variation of 20% was assessed (as in most studies) [6,21,23]. With an increase in the complications costs, direct treatment of patients become a more cost-effective strategy because more expensive complications are avoided. The Belgian costs included in this model being lower than those in the literature, the direct treatment strategy tends to be even more interesting.

As in the two previous sensitivity analyses, the medical care costs other than the therapy cost are always lower with the direct treatment strategy. The ICERs range between €22 540 and €23 552 for genotypes 1–4–5–6 and between €4 159 and €5 102 for genotypes 2–3 (Table 9).

Sensitivity analysis on the rates of SVR and EVR

With a variation of 20% of the SVR and EVR rates, the same conclusion can be drawn. Regardless of the level of SVR and EVR within the limits of the sensitivity analysis assessed, the costs of the medical care other than the therapy cost are lower with a direct treatment and the quality-adjusted life expectancy is better. The ICERs range between €21 095 and

Table 12 Sensitivity analysis on age

Genotypes 1–4–5–6	Minimal value	Baseline	Maximal value
First strategy			
Cost			
Drug costs	€15931.58	€15931.58	€15931.58
Medical care costs	€2393.64	€2113.92	€1187.47
Total costs	€18325.23	€18045.50	€17119.06
Efficacy			
Discounted quality-adjusted life years (QALY)	22.43	20.56	13.94
Second strategy			
Cost			
Drug costs	€10091.38	€9648.14	€7808.52
Medical care costs	€3052.27	€2735.50	€1666.09
Total costs	€13143.65	€12383.64	€9474.61
Efficacy			
Discounted QALY	22.16	20.32	13.67
Incremental cost-efficacy			
Cost/QALY	€18983.25	€23046.07	€28545.57
Genotypes 2–3			
First strategy			
Cost			
Drug costs	€8427.08	€8427.08	€8427.08
Medical care costs	€856.72	€756.60	€425.01
Total costs	€9283.8	€9183.68	€8852.09
Efficacy			
Discounted QALY	23.15	21.20	14.41
Second strategy			
Cost			
Drug costs	€5337.88	€5103.43	€4130.35
Medical care costs	€2044.92	€1877.96	€1288.46
Total costs	€7382.79	€6981.39	€5418.81
Efficacy			
Discounted QALY	22.63	20.73	13.88
Incremental cost-efficacy			
Cost/QALY	€3650.50	€4630.56	€6472.50

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Table 13 Sensitivity analysis on discount rate

Discount rate	ICER for genotypes 1–4–5–6	ICER for genotypes 2–3
Baseline: Benefits, 1.5%; Costs, 3%	€23046.07	€4630.56
Benefits, 0%; Costs, 3%	€18863.59	€3907.23
Benefits, 0%; Costs, 5%	€24111.63	€5623.40
Benefits, 0%; Costs, 0%	€7334.68	€110.88
Benefits, 3%; Costs, 3%	€27813.90	€5406.16
Benefits, 5%; Costs, 5%	€44942.27	€9371.79

ICER, incremental cost-effectiveness ratio.

€26 090 for genotypes 1–4–5–6 and between €3 453 and €6 414 for genotypes 2–3 (Table 10).

Sensitivity analysis on the duration of the study

For this analysis, a variation of 10 years was assessed. Results show that the benefits of a direct treatment increase with time. The ICERs range between €20 343 and €32 161 for genotypes 1–4–5–6 and between €3 988 and €6 806 for genotypes 2–3 (Table 11).

Sensitivity analysis on age at the beginning of the study

In this analysis, patients aged of 25 and 65 years at the beginning of the study were tested. Results show that the benefits of a direct treatment decrease with the age. The ICERs range between €18 983 and €28 546 for genotypes 1–4–5–6 and between €3 651 and €6 473 for genotypes 2–3 (Table 12).

Sensitivity analysis on discount rate

Belgian guidelines recommend testing different scenarios concerning the discount rate in the sensitivity analysis: 0% for benefits and 3% for costs, 0% for benefits and 5% for costs and finally 0%, 3% or 5% for both benefits and costs [27]. The ICERs range between €7 335 and €44 942 for genotypes 1–4–5–6 and between €111 and €9 372 for genotypes 2–3 (Table 13).

Probabilistic sensitivity analysis

A risk analysis on the uncertain parameters of the model with a Monte Carlo simulation was done using the software @risk 4.5 (Palisade Europe UK Ltd, London). Triangular distributions were introduced for each utilities using the ranges tested above. For SVR and EVR rates, beta distributions instead of normal distributions were used because standard errors were not available. Costs were modelled as a lognormal distribution due to the positive nature and positive skew of costs. After 1000 iterations, the 95% confidence interval for the ICER is €3 882 to €42 392 for genotypes 1–4–5–6 and €797 to €7 881 for genotypes 2–3 (Figs 2 & 3). By assuming a societal willingness-to-pay equal to €50 000 per QALY, 98.1% of ICERs fall below the limit for genotypes 1–4–5–6 and 100% for genotypes 2–3. If the limit is reduced to €20 000 per QALY, the percentage of accepted ICERs falls to 83.9% for genotypes 1–4–5–6 and to 99.6% for genotypes 2–3.

DISCUSSION

The cost-effectiveness of peg-interferon compared with interferon antiviral therapy for patients with mild chronic hepatitis C has already been addressed [22] but until recently, no studies addressed the cost-effectiveness of the treatment of those patients with peg-interferon and ribavirin *vs* follow-up of these patients and decision to treat when they progressed to a moderate stage. A recent study analysed this question and it was concluded that treating patients at a mild rather than at a moderate stage was generally cost-effective [28].

The aim of this study was to evaluate, in terms of cost and health, the direct treatment of patients with mild chronic hepatitis C (first strategy), in comparison with a simple monitoring of these patients and a treatment only once the stage of moderate chronic hepatitis was reached (second strategy). The therapy studied was the combination of pegylated interferon with ribavirin, i.e. the current standard therapy. Results show that a direct treatment is cost-effective for all genotypes and mainly for genotypes 2 and 3 at a threshold of €50 000/QALY (€23 046/QALY for genotypes 1–4–5–6 and €4 631/QALY for genotypes 2–3). The sensitivity analysis confirms these results and shows that only extreme circumstances relating to utilities for genotypes 1–4–5–6 lead to an unacceptable ICER for the society (>€50 000). A Monte Carlo simulation shows that the 95% confidence interval of the ICER is €3 882 to €42 392 for genotypes 1–4–5–6 and €797 to €7 881 for genotypes 2–3.

Even if an ICER less than €50 000/QALY is the generally accepted threshold representing good value for money in Western Europe [29–31], some countries like the Netherlands limit the threshold to €20 000/QALY [6]. In our model, after 1000 iterations, 98.1% of the ICERs fall below €50 000/QALY and 83.9% fall below €20 000/QALY for genotypes 1–4–5–6. For genotypes 2–3, 100% of the ICERs fall below €50 000/QALY and 99.6% fall bellow €20 000/QALY.

As in other studies [6,17–26], the SVR rates selected came from randomized clinical trials. Thus, these rates measured the efficacy and not the effectiveness of the therapy, i.e. the theoretical effect of the treatment and not its effect in the daily practice. In the 'real world', the SVR rate could be lower than the rate assessed in the model. In practice, a direct treatment would not be any more cost-effective if the

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

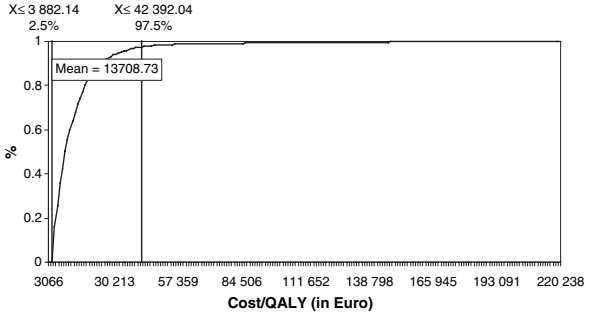

Fig. 2 Cumulative distribution of the ICER for genotypes 1–4–5–6.

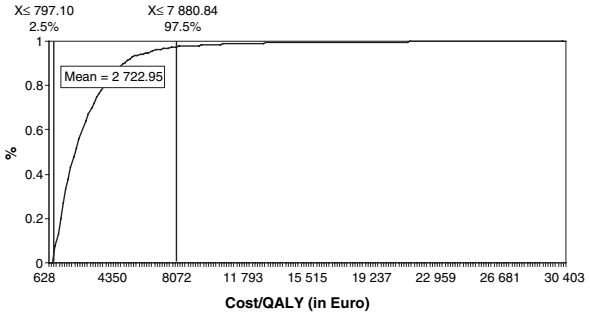

Fig. 3 Cumulative distribution of the ICER for genotypes 2–3.

SVR rates were lower than 25% for genotypes 1–4–5–6 and lower than 12% for genotypes 2 and 3.

In this study, patients were aged 45. However, by selecting older patients and thus patients with a lower life expectancy, the difference of QALY between the two strategies would decrease and the ICERs increase. Consequently, for genotypes 1–4–5–6, a direct treatment of patients older than 77 years would not be any more cost-effective (>€50 000/QALY). Conversely, decision to treat younger patients with mild hepatitis lead to a better ICER, as shown in the sensitivity analysis.

It would be interesting to study populations of alcoholic patients or patients co-infected by HIV or HBV, their rate of fibrosis progression being higher [32]. Sensitivity analysis shows that a faster progression supports a direct treatment but the parameters of the model being different for these patients, a specific analysis should be carried out.

In this study, we have adopted a conservative attitude and analyses of utilities show that any derogation to this conservative approach will support the cost effectiveness of direct treatment.

In the model, patients with hepatocellular carcinoma and compensated cirrhosis were supposed not to evolve towards liver transplantation, which was a conservative assumption. In practice, patients with hepatocellular carcinoma or even compensated cirrhosis are transplanted even if the number remains limited.

Some studies detailed the states of diuretic-sensitive ascites, refractory ascites, variceal haemorrhage and hepatic encephalopathy while others regrouped these states under the state 'decompensated cirrhosis'. Because guidelines on models [33–34] advise to use a simple structure and to avoid unnecessary complexities, we chose to regroup these states under the state of decompensated cirrhosis.

It should also be noted that extra-hepatic consequences of the disease were not taken into account in the study. A direct treatment (strategy 1) allows curing more patients and therefore taking into account these extra-hepatic consequences would even have benefited to this strategy. The approach adopted was thus conservative.

Moreover, the costs considered originated from an analysis based on the actual Belgian prices. Those being lower than prices imposed abroad, the approach adopted was again conservative. Indeed, as shown in the sensitivity analysis, higher costs result in a better ICER, more expensive complications being avoided. It was also supposed that the

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

level of care and thus the costs were similar for all chronic hepatitis C irrespective of the severity of the disease.

The conservative approach of using the most elevated utilities found in the literature [22] was also chosen. With lower utilities [18–21], the ICER would be improved, as shown in the sensitivity analysis. It should be noticed that results showed a large sensitivity associated with utility values.

A literature research concerning annual transition probabilities was performed. Similar values between later disease states, such as compensated and decompensated, were generally found [5,6,10,18,19,22,24,28] and thus were used in this study. However, there was less agreement about rates between mild, moderate and severe hepatitis C states. For these states, rates were calculated from a very recent study addressing this question [7] and large ranges were tested in the sensitivity analysis.

The technical validity of the model was checked by means of an extensive sensitivity analysis with extreme values as SVR rates equal to 0.

In conclusion, even if a direct treatment of people affected by mild chronic hepatitis C is more expensive, the complications are fewer and a greater number of people are cured, implying that such strategy is generally cost-effective at a threshold of €50 000/QALY.

ACKNOWLEDGEMENTS

No funding was received for the study. Logistical support was provided by Université catholique de Louvain, Cliniques universitaires Saint-Luc. Ralph Crott (health economic expert of the Belgian Health Care Knowledge Centre in Brussels) contributed to the probabilistic sensitivity analysis.

The role of Roche employees as co-authors pertained to approval of the manuscript.

REFERENCES

1 Shepherd J, Waugh N, Hewitson P. Combination therapy (interferon alfa and ribavirin) in the treatment of chronic hepatitis C: a rapid and systematic review. *Health Technol Assess* 2000; 4(33): 1–67.
2 Van Damme P, Thyssen A, Van Loock F. Epidemiology of hepatitis C in Belgium: present and future. *Acta Gastroenterol Belg* 2002; 65(2): 78–79.
3 Renard F, Autier M, Doumont D. L'hépatite C en Belgique, comment améliorer le dépistage et la prévention? Belgique: Ecole de santé Publique, Unité d'Education pour la santé, Série de dossiers techniques CFB, Report No. 05-34, 27 pp.
4 Delwaide J. Economic evaluation of chronic hepatitis C treatment by interferon-ribavirin combination therapy in Belgium. *Acta Gastroenterol Belg* 2002; 65(4): 233–236.
5 Sullivan SD, Jensen DM, Bernstein DE *et al.* Cost-effectiveness of combination peginterferon alpha-2a and ribavirin compared with interferon alpha-2b and ribavirin in patients with chronic hepatitis C. *Am J Gastroenterol* 2004; 99(8): 1490–1496.
6 Annemans L, Warie H, Nechelput M, Peraux B. A health economic model to assess the long term effects and cost-effectiveness of PEG IFN alpha-2a in hepatitis C virus infected patients. *Acta Gastroenterol Belg* 2004; 67(1): 1–8.
7 Alberti A. What is 'mild' chronic hepatitis C? *Hepatol Rev* 2005; 2: 19–23.
8 Salomon JA, Weinstein MC, Hammitt JK, Goldie SJ. Empirically calibrated model of hepatitis C virus infection in the United States. *Am J Epidemiol* 2002; 156(8): 761–773.
9 Michielsen P, Brenard R, Bourgeois N *et al.* Hepatitis C: screening, treatment and prevention practical guidelines. *Acta Gastroenterol Belg* 2003; 66(1): 15–19.
10 Wong JB, Koff RS. Watchful waiting with periodic liver biopsy versus immediate empirical therapy for histologically mild chronic hepatitis C. A cost-effectiveness analysis. *Ann Intern Med* 2000; 133(9): 665–675.
11 Levine RA. Treating histologically mild chronic hepatitis C: monotherapy, combination therapy, or tincture of time? *Ann Intern Med* 1998; 129(4): 323–326.
12 Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes, 2nd edn. Oxford: Oxford University Press, 1997.
13 Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Med Decis Making* 1993; 13(4): 322–338.
14 Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics* 1998; 13(4): 397–409.
15 Fried MW, Shiffman ML, Reddy KR *et al.* Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N Engl J Med* 2002; 347(13): 975–982.
16 Hadziyannis SJ, Sette H Jr, Morgan TR *et al.* Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. *Ann Intern Med* 2004; 140(5): 346–355.
17 Salomon JA, Weinstein MC, Hammitt JK, Goldie SJ. Cost-effectiveness of treatment for chronic hepatitis C infection in an evolving patient population. *JAMA* 2003; 290(2): 228–237.
18 Younossi ZM, Singer ME, McHutchison JG, Shermock KM. Cost effectiveness of interferon alpha2b combined with ribavirin for the treatment of chronic hepatitis C. *Hepatology* 1999; 30(5): 1318–1324.
19 Bennett WG, Inoue Y, Beck JR, Wong JB, Pauker SG, Davis GL. Estimates of the cost-effectiveness of a single course of interferon-alpha 2b in patients with histologically mild chronic hepatitis C. *Ann Intern Med* 1997; 127(10): 855–865.
20 Sagmeister M, Wong JB, Mullhaupt B, Renner EL. A pragmatic and cost-effective strategy of a combination therapy of interferon alpha-2b and ribavirin for the treatment of chronic hepatitis C. *Eur J Gastroenterol Hepatol* 2001; 13(5): 483–488.
21 Sennfalt K, Reichard O, Hultkrantz R, Wong JB, Jonsson D. Cost-effectiveness of interferon alfa-2b with and without ribavirin as therapy for chronic hepatitis C in Sweden. *Scand J Gastroenterol* 2001; 36(8): 870–876.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

22 Siebert U, Sroczynski G, Rossol S *et al.* Cost effectiveness of peginterferon alpha-2b plus ribavirin versus interferon alpha-2b plus ribavirin for initial treatment of chronic hepatitis C. *Gut* 2003; 52(3): 425–432.
23 Kim WR, Poterucha JJ, Hermans JE *et al.* Cost-effectiveness of 6 and 12 months of interferon-alpha therapy for chronic hepatitis C. *Ann Intern Med* 1997; 127(10): 866–874.
24 Wong JB, Poynard T, Ling MH, Albrecht JK, Pauker SG. Cost-effectiveness of 24 or 48 weeks of interferon alpha-2b alone or with ribavirin as initial treatment of chronic hepatitis C. International Hepatitis Interventional Therapy Group. *Am J Gastroenterol* 2000; 95(6): 1524–1530.
25 Jusot JF, Colin C. Cost-effectiveness analysis of strategies for hepatitis C screening in French blood recipients. *Eur J Public Health* 2001; 11(4): 373–379.
26 Wong JB, Nevens F. Cost-effectiveness of peginterferon alfa-2b plus ribavirin compared to interferon alfa-2b plus ribavirin as initial treatment of chronic hepatitis C in Belgium. *Acta Gastroenterol Belg* 2002; 65(2): 110–111.
27 Cleemput I, Crott R, Vrijens F, Huybrechts M, Van Wilder P, Ramaekers D. Recommandations provisoires pour les évaluations pharmacoéconomiques en Belgique. Bruxelles: Centre fédéral d'expertise des soins de santé (KCE), 2006; 6 A.D. May. Report No.: KCE Reports vol. 28B. Ref. D/2006/10.273/11.7.
28 Grieve R, Roberts J, Wright M *et al.* Cost effectiveness of interferon alpha or peginterferon alpha with ribavirin for histologically mild chronic hepatitis C. *Gut* 2006; 55(9): 1332–1338.
29 Shepard DS. Cost-effectiveness. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. Health and Medicine. New York: Oxford University Press, 1996. *J Ment Health Policy Econ* 1999; 2(2): 91–92.
30 Blackmore CC, Ramsey SD, Mann FA, Deyo RA. Cervical spine screening with CT in trauma patients: a cost-effectiveness analysis. *Radiology* 1999; 212(1): 117–125.
31 Roze S, Valentine WJ, Evers T, Palmer AJ. Acarbose in addition to existing treatments in patients with type 2 diabetes: health economic analysis in a German setting. *Curr Med Res Opin* 2006; 22(7): 1415–1424.
32 de Torres M, Poynard T. Risk factors for liver fibrosis progression in patients with chronic hepatitis C. *Ann Hepatol* 2003; 2(1): 5–11.
33 McCabe C, Dixon S. Testing the validity of cost-effectiveness models. *Pharmacoeconomics* 2000; 17(5): 501–513.
34 Hammerschmidt T, Goertz A, Wagenpfeil S, Neiss A, Wutzler P, Banz K. Validation of health economic models: the example of EVITA. *Value Health* 2003; 6(5): 551–559.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

3 A HEALTH ECONOMIC MODEL TO ASSESS THE COST-EFFECTIVENESS OF PEGYLATED INTERFERON α-2A AND RIBAVIRIN IN PATIENTS WITH MODERATE CHRONIC HEPATITIS C AND PERSISTENTLY NORMAL ALANINE AMINOTRANSFERASE LEVELS

S Gerkens, M Nechelput, L Annemans, B Peraux, C Beguin, Y Horsmans. A health economic model to assess the cost-effectiveness of pegylated interferon α-2a and ribavirin in patients with moderate chronic hepatitis C and persistently normal alanine aminotransferase levels. Acta gastroenterological belgica 2007 Apr-Jun;70(2):177-87.

NB: The definitive version is available at www.acta-gastroenterologica.be
© Universa Press.

A health economic model to assess the cost-effectiveness of pegylated interferon α-2a and ribavirin in patients with moderate chronic hepatitis C and persistently normal alanine aminotransferase levels

Sophie Gerkens[1], Myriam Nechelput[2], Lieven Annemans[3], Bénédict Peraux[2], Claire Beguin[1], Yves Horsmans[1]

(1) Université catholique de Louvain, Cliniques universitaires Saint-Luc, Bruxelles ; (2) NV ROCHE SA Bruxelles ; (3) Universiteit Gent, Vrije Universiteit Brussel.

Abstract

Background and study aims : **The treatment of patients with moderate chronic hepatitis C and persistently normal alanine aminotransferase levels is still under discussion and the cost-effectiveness of such strategy is unknown. The objective of this study was to estimate the cost-effectiveness of their treatment in comparison with no treatment.**

Patients and methods : **The assessed treatment is composed of pegylated interferon α-2a and ribavirin, which is the current standard treatment. Two groups were studied : patients with genotype 1 and patients with genotypes 2-3. At the beginning of the study, patients were aged of 45.**

Long-term economic and clinical outcomes over a 30 year period were predicted using a Markov simulation model. A health care payer perspective was chosen. Data were obtained from published literature. Variations of uncertainty parameters were assessed through a sensitivity analysis.

Results : **The incremental cost-effectiveness ratios (ICERs) were € 5,338/QALY for genotype 1 and € 1,080/QALY for genotypes 2-3. In the sensitivity analysis, ratios remained lower than € 20,000. A Monte Carlo simulation with 1,000 iterations gives a 95% confidence interval for the ICER of € 3,199 to € 8,972 for genotype 1 and € 56 to € 1,981 for genotypes 2-3.**

Conclusion : **Even though the treatment of these patients generates a cost, it has the advantage that in comparison with no treatment, a great number of people are cured, complications are less frequent and patients gain more quality-adjusted life-year (QALY), which involves an ICER considered as acceptable for the European society (< € 20,000).** (Acta gastroenterol. belg., **2007**, 70, **177-187**).

Key words : economic, cost-effectiveness, moderate chronic hepatitis C, persistently normal alanine aminotransferase, pegylated interferon, Markov model.

Introduction

Hepatitis C virus (HCV) is a major public health problem. Worldwide, it is estimated that 170 million persons are infected, i.e. 3% of the world population (1). In Belgium, the prevalence is estimated at nearly 1% (2). People infected by HCV can remain asymptomatic during more than 20 years and 35 to 80% of infected people are unaware of their seropositivity. Moreover, when the infection is diagnosed, few patients are treated (3). However, 80% of the infected persons will develop chronic hepatitis C, which will evolve for 20% of them to cirrhosis 20 to 30 years later. Afterwards, 3.9% of patients per year will develop a decompensated cirrhosis and 1.4% per year a hepatocellular carcinoma (4). In Belgium, more than 29% of the liver transplantations are related to hepatitis C (5).

Chronic hepatitis C can be classified into three different categories, according to the type and the amount of fibrosis. Mild chronic hepatitis C is characterized by no or few fibrosis, located in the portal area and without septa (METAVIR fibrosis scores of F0-F1). Moderate chronic hepatitis C shows portal or periportal fibrosis with some septa (METAVIR fibrosis score of F2) and severe chronic hepatitis C is determined by the presence of septal fibrosis without cirrhosis (METAVIR fibrosis score of F3) (6-8). Among patients with chronic hepatitis C, approximately 25% have persistently normal aminotransferase (PNALT) levels (9). PNALT levels is defined as normal serum alanine transaminase (ALT) over a 6-18 months period on three separate occasions with a minimum of 4 weeks apart (10).

At this time, no vaccine is available but curing the disease remains possible. The current standard therapy for patients with moderate to severe hepatitis C (METAVIR fibrosis scores of F2 and F3) and elevated ALT is the combination of pegylated interferon with ribavirin (1,11). The treatment of patients with persistently normal alanine aminotransferase is still under discussion (12). Indeed, these patients will not always develop a complication and treating them is stated to be expensive (1). However, it would be interesting to see whether, taking into account the long term complications, it would not be more effective and possibly less expensive to treat them.

The goal of this study was to compare two kinds of strategies :

- To treat patients with moderate chronic hepatitis C (METAVIR fibrosis scores of F2) and PNALT
- Not to treat and to only monitor patients with moderate chronic hepatitis C (METAVIR fibrosis scores of F2) and PNALT

With the growing cost of health care, efficacy and safety become insufficient for a good decision making.

Address for correspondence and for reprints : Yves Horsmans, Gastroenterology department, Cliniques universitaires Saint-Luc, Université catholique de Louvain, 10 Avenue Hippocrate box 2822, B-1200 Brussels, Belgium. E-mail : horsmans@gaen.ucl.ac.be.

Submission date : 24.10.2006
Acceptance date : 25.01.2007

In the current environment, the priority becomes to reduce the costs without deteriorating the quality of care, or to improve quality of care at a reasonable cost (13). Consequently, the effectiveness but also the costs of these two strategies were investigated in this study.

Methods and design

The Markov model

The cost-effectiveness was conducted using a Markov simulation model representing the natural history of chronic hepatitis C. This technique allows to model the evolution of a cohort of patients over time and assumes that individuals are always in one of a finite number of states of health, called "Markov state". They progress in a fixed unit of time, referred to as "Markov cycle". For each Markov state, a fixed annual transition probability towards the following state, a cost and a utility are referred (14,15). Annemans *et al.* recently developed a model for patients with chronic hepatitis C based on the literature (16). Within the framework of this study, this model was used and adapted in order to distinguish the various states of chronic hepatitis C, i.e. the scores of fibrosis.

The cycles considered in this study were annual, so each year, a patient evolved or not to the following state. At the beginning of the study, all patients had moderate chronic hepatitis C (METAVIR fibrosis score of F2) with PNALT and were 45 years old. Because in Belgium there was no statistical significant difference in HCV seroprevalence for gender (2), 50% of the studied population were men at the beginning of the study. Modifications of the age and of the proportion of men at the beginning of the study were tested in the sensitivity analysis.

This model assessed two strategies : to treat these patients or to not treat and to only monitor them. The programming was done in Microsoft excel.

For genotype 1, the treatment assessed was PEG IFN α-2a 180 mg once a week in combination with ribavirin 1000-1200 mg/day for 48 weeks, as recommended in the guidelines for patients with elevated ALT. For these genotypes, an early predictive test is performed after 12 weeks of treatment and if the viral load is not reduced by more than 2log10, the treatment is stopped (11). This early virological response (EVR) is determined by a quantitative test (Cobas Amplicor HCV Monitor test) which measures the reduction of HCV RNA and is positive when the test shows a reduction of more than 2 log10 or unquantifiable HCV RNA.

In the analysis, the duration of treatment considered if no EVR was detected was 16 weeks and not 12, in order to take into account the time to obtain the results and to stop the therapy.

For genotypes 2 and 3, the treatment considered was PEG IFN α-2a 180 mg once a week in combination with ribavirin 800 mg/day for 24 weeks. For these genotypes, no early predictive test is performed (11).

For each genotype, the sustained viral response rate (SVR) was calculated 24 weeks post treatment cessation and was defined as a polymerase chain reaction (PCR) negative test (absence of HCV RNA ; Cobas Amplicor HCV test). Patients with SVR were cured while the other patients could evolve to compensated cirrhosis. Afterwards, they could progress towards decompensated cirrhosis or hepatocellular carcinoma. Patients with decompensated cirrhosis could die, have a hepatocellular carcinoma or undergo a liver transplantation. Finally, patients with hepatocellular carcinoma or liver transplantation could also die (Fig. 1). As in the model developed by Annemans *et al.* (16), it was assumed that patients with compensated cirrhosis could not directly progress towards the death state and that a patient could not be found in two different states within the same year.

The viral response rates (SVR and EVR)

For genotype 1, the sustained viral response rate of the model came from the study completed by Fried *et al.* (17). This randomized clinical trial compared two types of treatments : PEG IFN α-2a 180 mg once a week in combination with ribavirin 1000-1200 mg/day for 48 weeks and Interferon α-2b 3 MIU three times per week in combination with ribavirin 1000-1200 mg/day for 48 weeks. Among the 298 patients with genotype 1 receiving PEG IFN α-2a and ribavirin, 138 had a SVR 24 weeks after the end of the treatment, which gives a SVR rate of 46.3%. In this study, patients had elevated ALT. Zeuzem *et al.* (10) assessed the SVR rate for patients with genotype 1 and PNALT. However, the treatment studied was PEG IFN α-2a 180 mg once a week in combination with ribavirin 800 mg/day and not 1000-1200 mg/day for 48 weeks. They found a SVR rate of 40.4% and concluded that this rate was similar to the rate for patients with elevated ALT. Hadziyannis *et al.* (18) tested different levels of ribavirin and concluded that ribavirin 1000-1200 mg/day for 48 weeks resulted in a significantly higher SVR rate than for ribavirin 800 mg for 48 weeks. Thus, the use of the SVR rate determined by Fried *et al.* (17) was the most suitable rate for this study because 1000-1200 mg/day of ribavirin was assessed. A variation of this rate was done in the sensitivity analysis. An EVR rate of 80.9% was determined for genotype 1 patients with PNALT and treated with peginterferon and ribavirin during 48 weeks (19).

For genotypes 2-3, SVR came from the study of Zeuzem *et al.* (10). This randomized clinical trial compared three groups of patients with PNALT : patients treated with PEG IFN α-2a 180 mg once a week in combination with ribavirin 800 mg/day for 48 weeks, patients treated with PEG IFN α-2a 180 mg once a week in combination with ribavirin 800 mg/day for 24 weeks and patients not treated. Patients were monitored for 72 weeks. Among the 58 patients with genotypes 2-3

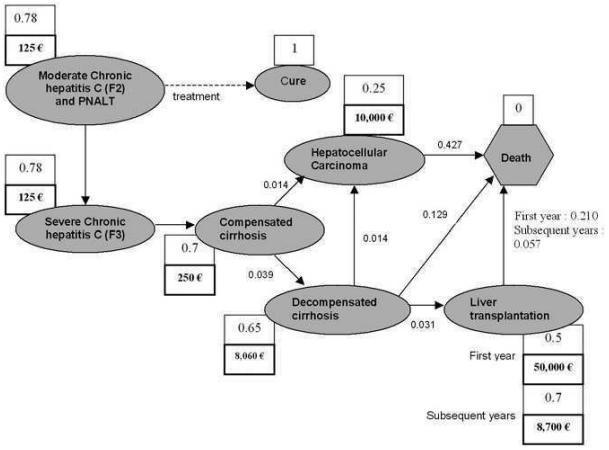

Fig. 1. — Markov model of HCV disease progression.
For each Markov state, a fixed annual transition probability towards the following state (xx), a cost (**xx**) and a utility (xx) are specified. Annual transition probabilities between the states of moderate chronic hepatitis C, severe chronic hepatitis C and compensated cirrhosis differ according to the patients age and gender (detailed in Table 1).

Table 1. — **Annual transition probabilities**

Markov States (references)	%
Fibrosis progression in men per age (20-25)	
40-49	0,0302 (0,0151-0,0532)
50-59	0,0700 (0,0409-0,0902)
60-69	0,1238 (0,0700-0,1954)
>= 70	0,1686 (0,0851-0,2677)
Fibrosis progression in women per age (20-25)	
40-49	0,0157 (0,0073-0,0325)
50-59	0,0364 (0,0157-0,0650)
60-69	0,0638 (0,0235-0,1322)
70-79	0,0862 (0,0454-0,1557)
>=80	0,1176 (0,0476-0,1988)
Compensated Cirrhosis to Decompensated Cirrhosis (4,27)	3.9 (2.0-8.3)
Compensated Cirrhosis to Hepatocellular carcinoma (4,28-29)	1.4 (1.0-4.4)
Decompensated Cirrhosis to Hepatocellular carcinoma (4,28-29)	1.4 (1.0-4.4)
Decompensated Cirrhosis to Liver transplantation (30)	3.1 (1.0-6.2)
Decompensated Cirrhosis to Death (4,30-31)	12.9 (6.5-19.3)
Hepatocellular Carcinoma to Death (4)	42.7 (33-86)
Liver transplantation to Death (First Year) (32-34)	21.0 (6-42)
Liver transplantation to Death (Subsequent Years) (32-34)	5.7 (2.4-11)

and treated with PEG IFN α-2a 180 mg once a week in combination with ribavirin 800 mg/day for 24 weeks, 42 patients had a SVR 24 weeks after the end of the treatment, which gives a SVR rate of 72.4%. In the untreated group, no patient cleared HCV RNA.

Annual transition probabilities

From previous studies (20-23), Salomon *et al.* (24) estimated annual transition probabilities between Metavir fibrosis score F2 to F4 for patients with elevated ALT. Results were stratified by age and gender. A longitudinal study performed by Hui *et al.* (25) compared the FPR for patients with PNALT and for patient with elevated ALT. The relative rate of fibrosis progression (FPR PNALT/FPR elevated ALT) was estimated from the figure representing the cumulative probability of fibrosis progression for patients with PNALT and with elevated ALT. This figure was analysed using a graphical analysis software called "Grab-It", which permitted to extract data from the figure. The median time (50th percentile) to progression for elevated ALT

occurred at year 7 and corresponded to the 28th percentile for PNALT. Consequently, at year 7 the relative rate of progression was 0.56 (0.28/0.50). By applying this rate to results of Salomon *et al.* (24), annual transition probabilities between Metavir fibrosis score F2 to F4 for patients with PNALT were obtained.

The other probabilities came from the model developed by Annemans *et al.* (16), themselves coming from the model of Younossi *et al.* (26) and based on previous studies (4,27-34) (Table 1). The mortality due to natural causes was also included in the model using age-specific mortality rates by gender (35).

Utilities

Utilities represent the preferences of individuals for a specific health state. They vary from 0, which represents death, to 1, which represents the perfect health state. These utilities are used to estimate the health-related quality of life and then the cost-effectiveness ratio in terms of cost per quality-adjusted life-year (QALY). For each strategy investigated, a QALY is calculated by multiplying the time spent in each health state by the utility of the state and then by summing the whole over the life expectancy (14). In other published models (26,30,36-37), several estimates of utilities can be found. In this study, the utilities relating to the states of moderate chronic hepatitis C, severe chronic hepatitis C and compensated cirrhosis came from the study of Bennett *et al.* (30) because fibrosis scores were distinguished. They used linear scaling and time trade-off methods. For the other states, the utilities of the model used by Annemans *et al.* (16) were preserved. Some of these utilities came from the study of Bennett *et al.* (30) and others came from the study of Kim *et al.* (31). Kim *et al.* (31) estimated utilities from a panel of hepatologists and a nurse specialist (Table 2). Deterioration of the quality of life due to the treatment was included in the model using the utility multiplier for peg-interferon and ribavirin assessed by Siebert *et al.* (38).

In order to take into account the differences between the studies, a sensitivity analysis, which assessed the range of values coming from the literature (16,26,30-31,36-37), was performed.

Complications costs

The Belgian health care payer perspective was adopted in this study. Various estimates of the costs are present in other published models (1,26,37,39-41).

For this study, the costs came from the model of Annemans *et al.* (16) because they were derived from a Belgian study performed by Wong in 2002 (41) (Table 3). The medical care costs of moderate and severe chronic hepatitis were supposed to be the same, these patients requiring no different caring.

Drug costs

The current public prices published by the "Institut national d'assurance maladie invalidité" (INAMI) in January 2006, VAT included and patient contribution deduced, were used. For genotype 1, the treatment cost € 18,187/patient. By taking into account the EVR rate, and thus the proportion of patient for which the treatment was stopped after 16 weeks, the mean treatment cost was reduced to € 15,865/patient. For genotypes 2-3, the treatment cost € 8,427/patient.

Table 2. — **Utilities estimates**

Utilities					
Markov State	Younossi 1999 (26)	Bennett 1997 (30)	Sagmeister 2001 (36)	Sennfält 2001 (37)	Model
utility multiplier for peg-interferon and ribavirin					0.90
SVR	1	1		1	1
Chronic hepatitis C	0.82 (0.6-0.9)		0.90		
Mild		0.82 (0.63-0.98)		0.82	0.82 (0.63-0.98)
Moderate to severe		0.78 (0.6-0.9)		0.78	0.78 (0.6-0.9)
Compensated cirrhosis	0.78 (0.5-0.9)	0.7 (0.5-0.9)	0.75	0.70	0.70 (0.5-0.9)
Decompensated cirrhosis	0.65 (0.3-0.88)				0.65 (0.3-0.88)
Hepatocellular carcinoma	0.25 (0.1-0.5)	0.1 (0.02-0.5)	0.20	0.10	0.25 (0.1-0.5)
Liver transplantation, first year	0.5 (0.11-0.70)	0.5 (0.11-0.70)	0.70	0.50	0.5 (0.11-0.70)
Liver transplantation, subsequent years	0.7 (0.24-0.87)	0.7 (0.24-0.87)	0.80	0.70	0.7 (0.24-0.87)

The utility multiplier used in the model came from the study of Siebert *et al.* (38). Other utilities came from the studies of Younossi *et al.* (26) and Bennett *et al.* (30). Younossi *et al.* (26) themselves based their estimates on the study of Bennett *et al.* (30) and Kim *et al.* (31). Bennett *et al.* (30) used linear scaling and time trade-off techniques to estimate utilities from a panel of hepatologists. In the study of Kim *et al.* (31), utilities were based on expert opinion (a panel of hepatologists and a nurse specialist).

Table 3. — **Costs estimates**

Costs per year						
Markov states	Wong 2000 (39)	Sennfält 2001 (37)	Jusot 2001 (40)	Younossi 1999 (26)	Shephred 2000 (1)	Wong 2002 (41)
Devise	$	$	€	$	£	€
Chronic hepatitis C			579	319	95	125
Compensated cirrhosis	110	163	1,400	425	237	250
Decompensated cirrhosis		6,885	11,500	22,100		8,060
Hepatocellular carcinoma			10,400		13,320	10,000
Liver transplantation						
First year	108,659	97,281	92,700	269,072		50,000
Subsequent years	18,976	16,093	7,900	25,910		8,700

Temporal characteristics of the study

Because the evolution of the disease is slow the period taken into account in this study was 30 years. Future costs were discounted at a rate of 3% and future utilities were discounted at a rate of 1.5%, as recommended by Belgian guidelines for cost-effectiveness analysis (42).

Sensitivity analysis

Considering the variations of input estimates from different sources, a sensitivity analysis was necessary. Firstly, an univariate sensitivity analysis assessing the impact of changes in key parameters on results was performed. Then, elasticity of each decision parameter was calculated to identify variables having an important impact on results. Finally, a probabilistic sensitivity analysis on all uncertain parameters using a Monte Carlo simulation with 1,000 iterations was done in order to determinate the 95% confidence interval of the ICERs.

Results

For genotype 1

The Markov technique had permitted to model the evolution of a cohort of patients with moderate chronic hepatitis C over a 30 year period. At each state of the model, a cost and a utility were associated and thus, the total cost and the quality-adjusted life-year (QALY) of the strategies could be calculated.

Treatment costs € 15,865 per patient. On the other hand, the studied model showed that the costs for medical care other than treatment cost were reduced by € 2,667 because 46.3% patients were cured after the treatment. In terms of health related life expectancy, people gained 2,47 QALY with a treatment.

Thus, even if the therapy was expensive, the other medical care costs were reduced and the average life expectancy in good health was higher, which resulted in an incremental cost-effectiveness ratio (ICER) of € 5,338/QALY (Table 4).

For genotypes 2-3

These patients were treated less longer and therefore the impact of the treatment cost (€ 8,427 per patient) was less important than with genotype 1. The costs for other medical care than therapy cost were 3.6 times lower thanks to the treatment (€ 1,589 versus € 5,760) and patients gained 3,94QALY. Consequently, the ICER was better than for genotype 1 and was € 1,080/QALY (Table 4).

Sensitivity analysis

The key parameters of the model were the annual transition probabilities, the utilities, the costs, the SVR rates, the duration of the study, the age at the beginning of the study, the discount rate, the proportion of men at the beginning of the study and the age-specific mortality rates by gender.

Univariate sensitivity analysis on the probabilities of evolution

The range of possible values of annual transition probabilities, estimated from the literature (see table 1), was assessed. With a faster progression of the disease, the ICER rate improved. In all situations considered, the medical care costs other than the therapy cost were lower with a treatment and the health-related life expectancy was better. The ICERs ranged between € 3,550 to € 6,411 for genotype 1 and between € 254 to € 1,693 for genotypes 2-3 (Table 5).

Univariate sensitivity analysis on the utilities

The range of possible utilities used by Younossi *et al.* (26) and Bennett *et al.* (30) was assessed for this sensitivity analysis (see table 2). The more the patients would consider their state as "unpleasant", the more the cost-effectiveness ratio improved. Conversely, if patients regarded their state as few disturbing, a treatment became less cost-effective. However, whatever the value

Table 4. — **Basecase results**

Genotype 1	Treatment	No treatment	Difference
Cost			
Drug cost	€ 15,865.03	€ 0.00	€ 15,865.03
Medical care cost	€ 3,092.64	€ 5,760.04	€ -2,667.40
Total cost	€ 18,957.67	€ 5,760.04	€ 13,197.63
Effectiveness			
Life expectancy (Years)	27.05	26.78	0.27
QALY	19.11	16.63	2.47
% cured people after 30 years	46.31%	0.00%	46.31%
Cost-effectiveness rate			
Cost/Life Year	€ 700.76	€ 215.10	€ 485.66
Cost/QALY	€ 992.20	€ 346.28	€ 645.93
ICER			€ 5,337.97/QALY
Genotypes 2-3	Treatment	No treatment	Difference
Cost			
Drug cost	€ 8,427.08	€ 0.00	€ 8,427.08
Medical care cost	€ 1,588.98	€ 5,760.04	€ -4,171.06
Total cost	€ 10,016.06	€ 5,760.04	€ 4,256.02
Effectiveness			
Life expectancy (Years)	27.21	26.78	0.43
QALY	20.57	16.63	3.94
% cured people after 30 years	72.41%	0.00%	72.41%
Cost-effectiveness rate			
Cost/Life Year	€ 368.14	€ 215.10	€ 153.04
Cost/QALY	€ 486.81	€ 346.28	€ 140.53
ICER			€ 1,080.06/QALY

of utility assessed, the health-related life expectancy remained higher with a treatment. The ICERs ranged between € 3,030 to € 11,689 for genotype 1 and between € 619 to € 2,305 for genotypes 2-3 (Table 5).

Univariate sensitivity analysis on the costs

For the sensitivity analysis on the costs, a variation of 20% was assessed (as in many studies) (16, 31, 37). With an increase in the complications costs, treatment of patients became more cost-effective because more expensive complications were avoided. As in the two preceding sensitivity analyses, the medical care costs other than the therapy cost were always lower with the treatment strategy. The ICERs ranged between € 5,122 to € 5,554 for genotype 1 and between € 868 to € 1,292 for genotypes 2-3 (Table 5).

Univariate sensitivity analysis on the rates of SVR and EVR

Because no confidence interval was given in the study of Fried *et al.* (17), a variation of 20% of the SVR and EVR rates was done. The same conclusion could be drawn. Whatever the levels of SVR and EVR within the limits of the sensitivity analysis assessed, the costs of the medical care other than the therapy cost were lower with a treatment and the health-related life expectancy was better. The ICERs ranged between € 4,908 to € 5,990 for genotype 1 and between € 723 to € 1,618 for genotypes 2-3 (Table 5). Variations beyond 20% were also assessed and ICERs became > € 20,000 from SVR rates below 14.9% for genotype 1 and below 7.9% for genotypes 2-3.

Univariate sensitivity analysis on the duration of the study

For this analysis, a variation of 10 years was assessed and the result showed that the benefits of a treatment increased with time. The ICERs ranged between € 3,866 to € 8,692 for genotype 1 and between € 502 to € 2,362 for genotypes 2-3 (Table 5).

Univariate sensitivity analysis on the age at the beginning of the study

Different values were tested in this sensitivity analysis. As results, the ICERs became > € 20,000 from 78 years old for genotype 1 and from 88 years old for genotypes 2-3.

Table 5. — **Univariate sensitivity analysis**

ICER for genotype 1			
Variables	Minimal values	Base case	Maximal values
Annual transition probabilities (Ranges of table 1)	€ 6,411.35	€ 5,337.97	€ 3,550.11
Utilities (Ranges of table 2)	€ 3,029.76	€ 5,337.97	€ 11,689.41
Costs (+/- 20%)	€ 5,553.75	€ 5,337.97	€ 5,122.20
SVR and EVR rates (+/- 20%)	€ 5,989.63	€ 5,337.97	€ 4,908.19
Time frame (20-40 years)	€ 8,692.08	€ 5,337.97	€ 3,865.69
Discount rate (0-5%)	€ 3,536.49	€ 5,337.97	€ 9,127.76
% men (0-100%)	€ 5,702.62	€ 5,337.97	€ 4,965.61
Age-specific mortality rates by gender (+/- 20%)	€ 5,189.74	€ 5,337.97	€ 5,488.06
ICER for genotypes 2-3			
Variables	Minimal values	Base case	Maximal values
Annual transition probabilities (Ranges of table 1)	€ 1,693.00	€ 1,080.06	€ 253.98
Utilities (Ranges of table 2)	€ 619.00	€ 1,080.06	€ 2,304.58
Costs (+/- 20%)	€ 1,291.76	€ 1,080.06	€ 868.36
SVR rate (+/- 20%)	€ 1,617.98	€ 1,080.06	€ 722.66
Time frame (20-40 years)	€ 2,362.40	€ 1,080.06	€ 502.26
Discount rate (0-5%)	€ 188.68	€ 1,080.06	€ 2,222.87
% men (0-100%)	€ 1,341.97	€ 1,080.06	€ 825.54
Age-specific mortality rates by gender (+/- 20%)	€ 1,028.45	€ 1,080.06	€ 1,132.11

Table 6. — **Elasticity**

Parameters	Elasticity	
	Genotype 1	Genotypes 2-3
Utility related to moderate hepatitis C (METAVIR fibrosis score F2)	3.140	3.015
Utility related to severe hepatitis C (METAVIR fibrosis score F3)	0.917	0.897
SVR rate	-1.019	-1.655
EVR rate	0.743	/

Univariate sensitivity analysis on the discount rate

Belgian guidelines recommend to test different scenarios concerning the discount rate in the sensitivity analysis : 0% for benefits and 3% for costs, 0% for benefits and 5% for costs and finally 0%, 3% or 5% for both benefits and costs (42). As results, ICERs ranged between € 3,536 to € 9,128 for genotype 1 and between € 189 to € 2,223 for genotypes 2-3 (Table 5).

Univariate sensitivity analysis on the proportion of men

In the basecase, 50% of the studied population were men. In this sensitivity analysis, extreme values were tested, i.e. 0 and 100%. As results, ICERs ranged between € 4,966 to € 5,703 for genotype 1 and between € 826 to € 1,342 for genotypes 2-3 (Table 5).

Univariate sensitivity analysis on the age-specific mortality rates by gender

For this analysis, a variation of 20% was assessed. As results, ICERs ranged between € 5,190 to € 5,488 for genotype 1 and between € 1,028 to € 1,132 for genotypes 2-3 (Table 5).

Elasticity of the parameters

In order to assess which variables had an important impact on results, elasticity of each decision parameter was calculated. Elasticity can be obtained by divided the percentage change of the results by the percentage change of the parameter. The more the absolute value of the elasticity was elevated, the higher the impact of the parameter was on the ICER. In this study, elasticity was

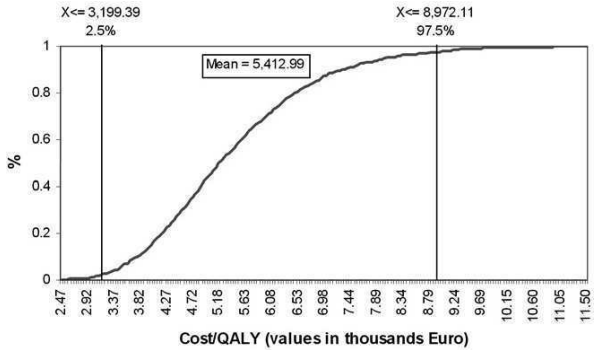

Fig. 2. — Cumulative distribution of the incremental cost-effectiveness ratio (ICER) for patients with genotype 1. These data are provided by a Monte Carlo simulation with 1,000 iterations and permit to determine the 95% confidence interval of the ICER.

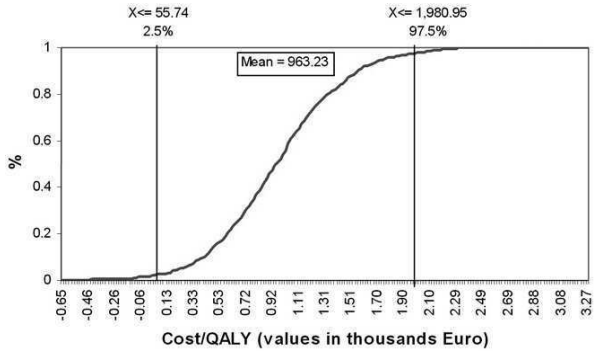

Fig. 3. — Cumulative distribution of the incremental cost-effectiveness ratio (ICER) for patients with genotypes 2 and 3. These data are provided by a Monte Carlo simulation with 1,000 iterations and permit to determine the 95% confidence interval of the ICER.

considered as elevated beyond a ratio of 0.5. As results, the utilities associated with the states of moderate and severe hepatitis C and the SVR and EVR rates had elevated elasticity (> 0.5) (Table 6).

Probabilistic sensitivity analysis

A risk analysis on the uncertain parameters of the model with a Monte Carlo simulation was done using the software @risk (Palisade Inc). Triangular distributions were introduced for each utility using the ranges tested above (see table 2). For the discount rate, a range of 0% to 5% was tested as advised in guidelines (42-43). For SVR and EVR rates, beta distributions instead of normal distributions were used because standard errors were not available. Costs were modelled as a lognormal distribution due to the positive nature and positive skew of costs. After 1,000 iterations, the 95% confidence interval for the ICER was € 3,199 to € 8,972 for genotype 1 and € 56 to € 1,981 for genotypes 2-3 (Figs. 2-3). By assuming a societal willingness-to-pay equal to € 20,000 per QALY, 100% of the ICERs fall bellow the limit for genotype 1 and for genotypes 2-3.

Discussion

In the literature, the decision to treat or not patients with PNALT is still under discussion (12,44,45). The aim of this study was to assess, in terms of cost and health, the impact of such decision, i.e. to determine the ICER of the treatment of patients with PNALT compared to no treatment. The therapy studied was the combination of pegylated interferon and ribavirin, i.e. the standard treatment for patients with elevated ALT (1,11).

Compared to no treatment, results showed that the treatment was a cost-effective strategy for all genotypes and mainly for genotypes 2 and 3 (€ 5,338/QALY for genotype 1 and € 1,080/QALY for genotypes 2-3). Indeed, in the Netherlands, a ratio below € 20,000/QALY is considered as cost-effective whereas, in US, the limit is $ 50,000/QALY (46-47). Our results were thus far from those exceeding the limit acceptable for the society. The sensitivity analysis confirmed these results and showed that utility associated to the state of moderate chronic hepatitis had the biggest impact on results. A Monte Carlo simulation showed that the 95% confidence interval was € 3,199 to € 8,972 for genotype 1 and € 56 to € 1,981 for genotypes 2-3.

As in other studies (16,24,26,30-31,36-41), the SVR rates selected came from randomized clinical trials. Thus, these rates measured the efficacy and not the effectiveness of the therapy, i.e. the theoretical effect of the treatment and not its effect in practice. In the "real world", the SVR rate is maybe lower than the rate assessed in the model. In practice, a treatment would not be any more cost-effective if the SVR rate went down below 14.9% for genotype 1 and below 7.9% for genotypes 2 and 3.

In this study, patients were aged of 45. However, by selecting older patients and thus patients with a lower life expectancy, the difference of quality-adjusted life-year between the two strategies was more limited and the cost-effectiveness ratio increased. Consequently, a treatment was not any more cost-effective (> € 20,000/QALY) from 78 years old for genotype 1 and from 88 years old for genotypes 2-3.

In other models of HCV disease progression, similar annual transition probabilities between later disease states, such as decompensated cirrhosis and liver transplantation, were generally found (16,26,30,38-39). However, there was less agreement about rates between Metavir fibrosis scores F2 to F4. In this model, these rates came from the study of Salomon *et al.* (24) because they were stratified by age and gender and large ranges were tested in the sensitivity analysis.

Annual transition probabilities between Metavir fibrosis scores F2 to F4 were adapted to patients with PNALT from only one study (25). Few studies have compared the progression rates of fibrosis between patients with PNALT and patients with elevated ALT and their methodology appears to be limited (25, 48-51). In the study of Hui *et al.* (25), ALT determination was only performed yearly. Moreover, the number of patients with F2 fibrosis score in this study was very small (9 in the group with PNALT and 11 in the group with elevated ALT). More data on the progression rates of fibrosis for patients with PNALT are thus needed. The impact of such limitation was reduced by performing a sensitivity analysis on these progression rates.

It would be interesting to develop a specific model on populations at risk like alcoholic patients or patients co-infected by HIV or HBV. For these patients, annual transition probabilities are higher (52). Moreover, for these patients, health care costs can be expected higher and utilities lower than for patients with hepatitis C alone. As shown in the sensitivity analysis, higher annual transition probabilities, higher health care costs and lower utilities were in favour of the treatment strategy. Not include these patients in this analysis was thus a conservative assumption.

In the model, patients with hepatocellular carcinoma and compensated cirrhosis were supposed not to evolve towards liver transplantation, which was a conservative assumption. In practice, patients with hepatocellular carcinoma or even compensated cirrhosis are transplanted even if their number remains limited.

It should also be noted that extra-hepatic consequences of the disease, as cryoglobulinemia and glomerulonephritis (53-55), were not taken into account in this study. However, taking into account these extra-hepatic consequences would increase health care costs for patients who are not cured and as shown in the sensitivity analysis, higher health care costs improved the ICERs. The approach adopted was thus conservative.

Moreover, the costs considered originated from an analysis based on the actual Belgian prices. Those being lower than the prices imposed abroad, the ICERs tends to be improved in other countries with similar mortality rates, annual transition probabilities and utilities. It was also assumed that the costs between the various degrees of severity of chronic hepatitis C were the same, their caring being not different.

In conclusion, even if the treatment of people with moderate chronic hepatitis C and PNALT generates a cost compared to no treatment, a great number of people are cured, complications are less frequent and patients gain more QALYs, which involves an ICER acceptable for the European society (< € 20,000).

Bibliography

1. SHEPHERD J., WAUGH N., HEWITSON P. Combination therapy (interferon alfa and ribavirin) in the treatment of chronic hepatitis C : a rapid and systematic review. *Health Technol. Assess.*, 2000, **4** : 1-67.
2. VAN DAMME P., THYSSEN A., VAN LOOCK F. Epidemiology of hepatitis C in Belgium : present and future. *Acta Gastroenterol. Belg.*, 2002, **65** : 78-79.
3. RENARD F., AUTIER M., DOUMONT D. L'hépatite C en Belgique, comment améliorer le dépistage et la prévention ? Belgique : Ecole de santé Publique, Unité d'Education pour la Santé, 2005. Report no. : 05-34.
4. FATTOVICH G., GIUSTINA G., DEGOS F., TREMOLADA F., DIODATI G., ALMASIO P., NEVENS F., SOLINAS A., MURA D., BROUWER J.T., THOMAS H., NJAPOUM C., CASARIN C., BONETTI P., FUSCHI P., BASHO J., TOCCO A., BHALLA A., GALASSINI R., NOVENTA F., SCHALM S.W., REALDI G. Morbidity and mortality in compensated cirrhosis type C : a retrospective follow-up study of 384 patients. *Gastroenterology*, 1997, **112** : 463-472.
5. DE HEMPTINNE B. On behalf of the Belgian liver transplantation activity in adults patients in Belgium. *Acta Gastroenterol. Belg.*, 2001, **64** : D04.
6. BEDOSSA P., POYNARD T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology*, 1996, **24** : 289-293.
7. ALBERTI A. What is "mild" chronic hepatitis C ? *Hepatology Rev.*, 2005, **2** : 19-23.

8. SALOMON J.A., WEINSTEIN M.C., HAMMITT J.K., GOLDIE S.J. Empirically calibrated model of hepatitis C virus infection in the United States. *Am. J. Epidemiol.*, 2002, **156** : 761-773.
9. OKANOUE T., MAKIYAMA A., NAKAYAMA M., SUMIDA Y., MITSUYOSHI H., NAKAJIMA T., YASUI K., MINAMI M., ITOH Y. A follow-up study to determine the value of liver biopsy and need for antiviral therapy for hepatitis C virus carriers with persistently normal serum aminotransferase. *J. Hepatol.*, 2005, **43** : 599-605.
10. ZEUZEM S., DIAGO M., GANE E., REDDY K.R., POCKROS P., PRATI D., SHIFFMAN M., FARCI P., GITLIN N., O'BRIEN C.B., LAMOUR F., LARDELLI P. Peginterferon alfa-2a (40 kilodaltons) and ribavirin in patients with chronic hepatitis C and normal aminotransferase levels. *Gastroenterology*, 2004, **127** : 1724-1732.
11. MICHIELSEN P., BRENARD R., BOURGEOIS N., DE GALOCSY C.H., DELWAIDE J., HENRION J., HORSMANS Y., NEVENS F., REYNAERT H., ROBAEYS G., SPRENGERS D., VAN VLIERBERGHE H. Hepatitis C : screening, treatment and prevention practical guidelines. *Acta Gastroenterol. Belg.*, 2003, **66** : 15-19.
12. SHIFFMAN M.L., DIAGO M., TRAN A., POCKROS P., REINDOLLAR R., PRATI D., RODRIGUEZ-TORRES M., LARDELLI P., BLOTNER S., ZEUZEM S. Chronic hepatitis C in patients with persistently normal alanine transaminase levels. *Clin. Gastroenterol. Hepatol.*, 2006, **4** : 645-652.
13. DRUMMOND MF., O'BRIEN BJ., STODDART GL., TORRANCE GW. Methods for the Economic Evaluation of Health Care Programmes. 2 ed. Oxford : Oxford University Press, 1997.
14. SONNENBERG FA., BECK JR. Markov models in medical decision making : a practical guide. *Med. Decis. Making*, 1993, **13** : 322-338.
15. BRIGGS A., SCULPHER M. An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics*, 1998, **13** : 397-409.
16. ANNEMANS L., WARIE H., NECHELPUT M., PERAUX B. A health economic model to assess the long term effects and cost-effectiveness of PEG IFN alpha-2a in hepatitis C virus infected patients. *Acta Gastroenterol. Belg.*, 2004, **67** : 1-8.
17. FRIED M.W., SHIFFMAN M.L., REDDY K.R., SMITH C., MARINOS G., GONCALES F.L. JR., HAUSSINGER D., DIAGO M., CAROSI G., DHUMEAUX D., CRAXI A., LIN A., HOFFMAN J., YU J. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. *N. Engl. J. Med.*, 2002, **347** : 975-982.
18. HADZIYANNIS S.J., SETTE H.J.R., MORGAN T.R., BALAN V., DIAGO M., MARCELLIN P., RAMADORI G., BODENHEIMER H. JR., BERNSTEIN D., RIZZETTO M., ZEUZEM S., POCKROS P.J., LIN A., ACKRILL A.M. Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C : a randomized study of treatment duration and ribavirin dose. *Ann. Intern. Med.*, 2004, **140** : 346-355.
19. POCKROS P.J. Early prediction of sustained virological response (SVR) during treatment with peginterferon alfa-2a (40KD) (PEGASYS®) plus ribavirin (RBV) (COPEGUS®) in patients with chronic hepatitis C (CHC) and persistently normal alanine aminotransferase (ALT) levels : Results of a multinational trial. *J. Hepatol.*, Suppl. 1, 2004 [abstract].
20. POYNARD T., BEDOSSA P., OPOLON P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. *Lancet*, 1997, **349** : 825-832.
21. SHIRATORI Y., IMAZEKI F., MORIYAMA M., YANO M., ARAKAWA Y., YOKOSUKA O., KUROKI T., NISHIGUCHI S., SATA M., YAMADA G., FUJIYAMA S., YOSHIDA H., OMATA M. Histologic improvement of fibrosis in patients with hepatitis C who have sustained response to interferon therapy. *Ann. Intern. Med.*, 2000, **132** : 517-524.
22. SOBESKY R., MATHURIN P., CHARLOTTE F., MOUSSALLI J., OLIVI M., VIDAUD M., RATZIU V., OPOLON P., POYNARD T. Modeling the impact of interferon alfa treatment on liver fibrosis progression in chronic hepatitis C : a dynamic view. The Multivirc Group. *Gastroenterology*, 1999, **116** : 378-386.
23. YANO M., YATSUHASHI H., INOUE O., INOKUCHI K., KOGA M. Epidemiology and long term prognosis of hepatitis C virus infection in Japan. *Gut*, 1993, **34** : S13-S16.
24. SALOMON J.A., WEINSTEIN M.C., HAMMITT J.K., GOLDIE S.J. Cost-effectiveness of treatment for chronic hepatitis C infection in an evolving patient population. *JAMA*, 2003, **290** : 228-237.
25. HUI C.K., BELAYE T., MONTEGRANDE K., WRIGHT T.L. A comparison in the progression of liver fibrosis in chronic hepatitis C between persistently normal and elevated transaminase. *J. Hepatol.*, 2003, **38** : 511-517.
26. YOUNOSSI Z.M., SINGER M.E., MC HUTCHISON J.G., SHERMOCK K.M. Cost effectiveness of interferon alpha2b combined with ribavirin for the treatment of chronic hepatitis C. *Hepatology*, 1999, **30** : 1318-1324.
27. GINES P., QUINTERO E., ARROYO V., TERES J., BRUGUERA M., RIMOLA A., CABALLERIA J., RODES J., ROZMAN C. Compensated cirrhosis : natural history and prognostic factors. *Hepatology*, 1987, **7** : 122-128.
28. TSUKUMA H., HIYAMA T., TANAKA S., NAKAO M., YABUUCHI T., KITAMURA T., NAKANISHI K., FUJIMOTO I., INOUE A., YAMAZAKI H., KAWASHIMA T. Risk factors for hepatocellular carcinoma among patients with chronic liver disease. *N. Engl. J. Med.*, 1993, **328** : 1797-1801.
29. COLOMBO M., DE FRANCHIS R., DEL NINNO E., SANGIOVANNI A., DE FAZIO C., TOMMASINI M., DONATO M.F., PIVA A., DI CARLO V., DIOGUARDI N. Hepatocellular carcinoma in Italian patients with cirrhosis. *N. Engl. J. Med.*, 1991, **325** : 675-680.
30. BENNETT W.G., INOUE Y., BECK J.R., WONG J.B., PAUKER S.G., DAVIS G.L. Estimates of the cost-effectiveness of a single course of interferon-alpha 2b in patients with histologically mild chronic hepatitis C. *Ann. Intern. Med.*, 1997, **127** : 855-865.
31. KIM W.R., POTERUCHA J.J., HERMANS J.E., THERNEAU T.M., DICKSON E.R., EVANS R.W., GROSS JB. Jr. Cost-effectiveness of 6 and 12 months of interferon-alpha therapy for chronic hepatitis C. *Ann. Intern. Med.*, 1997, **127** : 866-874.
32. ASCHER N.L., LAKE J.R., EMOND J., ROBERTS J. Liver transplantation for hepatitis C virus-related cirrhosis. *Hepatology*, 1994, **20** : 24S-27S.
33. DETRE K.M., BELLE S.H., LOMARDDERO M. Liver transplantation for chronic viral hepatitis. *Viral Hepatitis Rev.*, 1996, **2** : 219-228.
34. KILPE V.E., KRAKAUER H., WREN R.E. An analysis of liver transplant experience from 37 transplant centers as reported to Medicare. *Transplantation*, 1993, **56** : 554-561.
35. Service public fédéral Economie, PME, Classes moyennes et Energie : Démographie mathématique. Tables de mortalité 2004 et 2002-2004. Bruxelles : Service public fédéral Economie, PME, Classes moyennes et Energie. Direction générale Statistique et Information économique, 2006.
36. SAGMEISTER M., WONG J.B., MULLHAUPT B., RENNER E.L. A pragmatic and cost-effective strategy of a combination therapy of interferon alpha-2b and ribavirin for the treatment of chronic hepatitis C. *Eur. J. Gastroenterol. Hepatol.*, 2001, **13** : 483-488.
37. SENNFALT K., REICHARD O., HULTKRANTZ R., WONG J.B., JONSSON D. Cost-effectiveness of interferon alfa-2b with and without ribavirin as therapy for chronic hepatitis C in Sweden. *Scand. J. Gastroenterol.*, 2001, **36** : 870-876.
38. SIEBERT U., SROCZYNSKI G., ROSSOL S., WASEM J., RAVENS-SIEBERER U., KURTH B.M., MANNS M.P., MCHUTCHISON J.G., WONG J.B. ; GERMAN HEPATITIS C MODEL (GEHMO) GROUP, INTERNATIONAL HEPATITIS INTERVENTIONAL THERAPY (IHIT) GROUP. Cost effectiveness of peginterferon alpha-2b plus ribavirin versus interferon alpha-2b plus ribavirin for initial treatment of chronic hepatitis C. *Gut*, 2003, **52** : 425-432.
39. WONG J.B., POYNARD T., LING M.H., ALBRECHT J.K., PAUKER S.G. Cost-effectiveness of 24 or 48 weeks of interferon alpha-2b alone or with ribavirin as initial treatment of chronic hepatitis C. International Hepatitis Interventional Therapy Group. *Am. J. Gastroenterol.*, 2000, **95** : 1524-1530.
40. JUSOT J.F., COLIN C. Cost-effectiveness analysis of strategies for hepatitis C screening in French blood recipients. *Eur. J. Public Health*, 2001, **11** : 373-379.
41. WONG J.B., NEVENS F. Cost-effectiveness of peginterferon alfa-2b plus ribavirin compared to interferon alfa-2b plus ribavirin as initial treatment of chroni hepatitis C in Belgium. *Acta Gastroenterol. Belg.*, 2002, **65** : 110-111.
42. CLEEMPUT I., CROTT R., VRIJENS F., HUYBRECHTS M., VAN WILDER P., RAMAEKERS D. Recommandations provisoires pour les évaluations pharmacoéconomiques en Belgique. Bruxelles : Centre fédéral d'expertise des soins de santé (KCE), 2006. Report no. : D/2006/10.273/11.
43. MULLINS C.D., OGILVIE S. Emerging standardization in pharmacoeconomics. *Clin. Ther.*, 1998, **20** : 1194-1202.
44. ROBAEYS G., BUNTINX F., BOTTIEAU E., BOURGEOIS S., BRENARD R., COLLE I., DE BIE J., MATHEI C., MULKAY J.P., VAN DAMME P., VAN RANST M., VERRANDO R., MICHIELSEN P., BOURGEOIS N., BRENARD R., DE GALOCSY C.H., DELWAIDE J., HENRION J., HORSMANS Y., MICHIELSEN P., REYNAERT H., ROBAEYS G., SPRENGERS D. ; STEERING COMMITTEE OF THE BELGIAN ASSOCIATION FOR THE STUDY OF THE LIVER (BASL). Guidelines for the management of chronic hepatitis C in patients infected after substance use. *Acta Gastroenterol. Belg.*, 2005, **68** : 38-45.
45. VERSLYPE C., MICHIELSEN P., ADLER M., ORLENT H., SPRENGERS D., DELWAIDE J., D'HEYGERE F., LANGLET P., BRENARD R., COLLE I., REYNAERT H., STARKEL P., HENRION J. ; STEERING COMMITTEE OF THE BELGIAN ASSOCIATION FOR THE

STUDY OF THE LIVER. The management of patients with mild hepatitis C. *Acta Gastroenterol. Belg.*, 2005, **68** : 314-8.
46. MESSORI A., TRIPPOLI S., VAIANI M. Efficacy, safety, and cost of new anticancer drugs. Price needs to be evaluated against effectiveness. *BMJ*, 2002, **325** : 1302.
47. SIMOONS M.L. Cholesterol-lowering therapy, a recommendation from the Health Council. Ned Tijdschr Geneeskd, 2000, 144 : 2442-2444.
48. JAMAL M.M., SONI A., QUINN P.G., WHEELER D.E., ARORA S., JOHNSTON D.E. Clinical features of hepatitis C-infected patients with persistently normal alanine transaminase levels in the Southwestern United States. *Hepatology*, 1999, **30** : 1307-1311.
49. KYRLAGKITSIS I., PORTMANN B., SMITH H., O'GRADY J., CRAMP M.E. Liver histology and progression of fibrosis in individuals with chronic hepatitis C and persistently normal ALT. *Am. J. Gastroenterol.*, 2003, **98** : 1588-1593.
50. MATHURIN P., MOUSSALLI J., CADRANEL J.F., THIBAULT V., CHARLOTTE F., DUMOUCHEL P., CAZIER A., HURAUX J.M., DEVERGIE B., VIDAUD M., OPOLON P., POYNARD T. Slow progression rate of fibrosis in hepatitis C virus patients with persistently normal alanine transaminase activity. *Hepatology*, 1998, **27** : 868-872.
51. RENOU C., HALFON P., POL S., CACOUB P., JOUVE E., BRONOWICKI J.P., ARPURT J.P., RIFFLET H., PICON M., CAUSSE X., CANVA V., DENIS J., TRAN A., BOURLIERE M., OUZAN D., PARIENTE A., DANTIN S., ALRIC L., CARTIER V., REVILLE M., CAILLAT-ZUCMAN S. Histological features and HLA class II alleles in hepatitis C virus chronically infected patients with persistently normal alanine aminotransferase levels. *Gut*, 2002, **51** : 585-590.
52. DE TORRES M., POYNARD T. Risk factors for liver fibrosis progression in patients with chronic hepatitis C. *Ann. Hepatol.*, 2003, **2** : 5-11.
53. KOFF RS., DIENSTAG JL. Extrahepatic manifestations of hepatitis C and the association with alcoholic liver disease. *Semin. Liver Dis.*, 1995, **15** : 101-109.
54. LUNEL F., MUSSET L., CACOUB P., FRANGEUL L., CRESTA P., PERRIN M., GRIPPON P., HOANG C., VALLA D., PIETTE J.C., HURAUX J.M., OPOLON P. Cryoglobulinemia in chronic liver diseases : role of hepatitis C virus and liver damage. *Gastroenterology*, 1994, **106** : 1291-1300.
55. MARCELLIN P., BENHAMOU J.P. Autoimmune disorders associated with hepatitis C. *Prog. Liver Dis.*, 1995, **13** : 247-267.

4 CONCLUSION AND PERSPECTIVES

In this part, we used Markov models to determine the "cost-effectiveness" impact of the treatment of patients with moderate chronic hepatitis C and PNALT, and of the immediate treatment of patients with mild chronic hepatitis C and elevated ALT. To respond to such questions, the cascade of costs and effects were extrapolated in the long term, and allowed us to obtain incremental cost-effectiveness ratios (ICER). To handle uncertainty, a probabilistic sensitivity analysis was performed and allowed us to obtain the 95% confidence intervals of the ICER.

The analysis has shown that the treatment of patients with moderate chronic hepatitis C and PNALT should be promoted, especially for patients with genotypes 2-3. Indeed, the 95% CI of the ICER were €3,199/QALY - €8,972/QALY for genotype 1 and €56/QALY - €1,981/QALY for genotypes 2-3, which can be considered as acceptable to society. On the other hand, for patients with mild chronic hepatitis C and with genotypes 1-4-5-6, results were more moderate. The 95% CI of the ICER were €3,882-€42,392 for genotypes 1-4-5-6 and €797-€7,881 for genotypes 2-3.

The acceptability of the ICER would depend of the societal willingness to pay, which usually varies between €20,000/QALY and €50,000/QALY according to the country (1;2). Nevertheless, it should be noted that in Belgium, no explicit cost-effectiveness threshold is determined by the Drug Reimbursement Committee or the National Health Institute. The final decision to reimburse the treatment of these patients also takes into account other factors such as the burden of disease, or other clinical, ethical and political considerations.

Moreover, a comparison of results with other programs should be done. Indeed, the health care budget being limited, the reimbursement of one health care program threatens the reimbursement of other programs.

From a methodological point of view, we tested the descriptive validity, the technical validity and the convergent validity of the models.

The aim of the descriptive or face validity is to test if the structure of the model is acceptable and is a simplified but adequate representation of reality (3;4). In this thesis, the models were constructed in collaboration with a gastroenterologist and were based on the natural history of the disease. To avoid complexities, the states of ascites, variceal bleeding and hepatic encephalopathy were regrouped under the state of decompensated cirrhosis. Such a choice had no significant impact on results. Indeed, only the states at the beginning of the model showed high elasticity for the parameters.

The technical validity was also checked by testing the impact of extreme hypotheses on results, as for example SVR rate of 0 and 100% (3;4).

To test the convergent validity, we tested if our models gave similar results compared to other models constructed independently (3). Compared to other models on hepatitis C, our Belgian medical health care cost estimates were usually lower, which was a conservative approach, lower costs going in favor of the absence of treatment. As explained in the research questions, we also used the most elevated utilities for the model on the treatment of mild chronic hepatitis C, utilities having an important impact on results for this model. Then,

concerning transition probabilities, we used similar values than in other models (5-8). Moreover, we found two recent studies assessing the same subjects (9;10). One of the study also assessed the treatment of patients with chronic hepatitis C and PNALT with peginterferon and ribavirin and also found that such strategy was cost-effective compared to no treatment for patients with genotypes 2-3 and for patients with genotype 1 (10). The second study evaluated the direct treatment of patients with mild chronic hepatitis C and found, as in our study, that the direct treatment of genotype non-1 patients with pegylated interferon and ribavirin was a cost-effective strategy (9). On the other hand, they found that the direct treatment of patients with genotype 1 was not cost-effective. This may partially be explained by the fact that in our model, estimates were aggregated for genotypes 1-4-5-6. Future analyses for these patients' subgroups are thus needed in the future.

Moreover, the validation of these results in daily practice is needed. Even if we tested from which criteria the results would become ineffective, the predictive validity of the model would be interesting to test. The aim of the predictive validity is to compare final outcomes with prospective observational studies and to determine if the results of the model are similar to results in real life (3;4). However, such an approach is time consuming and is relatively expensive. So, it is preferable that health care decision makers do not wait for such a validation before to approve reimbursement. Indeed, the predictive validity has been judged to be valuable but not absolutely essential (11;12).

In the future, it would thus be interesting to develop a register of patients with chronic hepatitis C both to validate the predictability of

these models and to measure the health care quality in this field. It would also be interesting to use Belgian databases to test the predictive validity of modeling studies. However, currently, a long-term follow-up of patients with all hospitalizations and outpatients' data is difficult to obtain. Without an improved link between all existing databases, to assess the predictive validity through retrospective observational studies remains a difficult and even an impossible task. So, beside an improvement of the link between existing databases, including the obligation to hold registration of patients in the reimbursement's conditions would be valuable.

It should also be noted that pegylated interferon associated with ribavirin is the current standard treatment but in the future, when new therapies will be available, they could be compared using these models.

Concerning our models weaknesses, it should also be noted that since their publications, new Belgian health economic guidelines have been published in the literature (13). After an analysis of these guidelines, some points have to be highlighted.

Firstly, these guidelines advise to perform a systematic review of the clinical and economic literature on the studied topic. Nevertheless, even if we conducted a literature research via Pubmed using keywords such as chronic hepatitis c, cost* or econ* at the beginning of the study, this research should have been performed in a more systematic way as described in these guidelines, which would allow us to identify all available evidences. Moreover, even if the choice of our models' parameters was based both on a previous model (14) and on the literature research using a conservative approach (i.e. in favour of no immediate treatment in model 1 and of no treatment in

model 2), we did not assess the quality of every identified study with a quality assessment checklist. The non respect of these new guidelines is thus a study's weakness. Nevertheless, this weakness was reduced by the use of a conservative approach and by the use of a large probabilistic sensitivity analysis.

Secondly, we only took into account the national health institute perspective to assess the treatment cost. Out-of-pocket expenses for the patient were not included. By including these out-of-pocket expenses in the calculation, both the treatment cost and the ICER would be increased, but not in a significant way.

Thirdly, concerning the treatment analyzed in the models, the choice was based on the current standard of care as advised in the Belgium health economic guidelines (13), i.e. pegylated interferon and ribavirin. Some other models have also analyzed the association of standard interferon and ribavirin. However, even if standard interferon and ribavirin cost less, such a choice would be justified neither by Belgian clinical guidelines (15) nor by Belgian health economic guidelines (13), pegylated interferon associated with ribavirin having been proved to be the most effective treatment for chronic hepatitis C (15-18). Then, in the second model, the choice of no treatment as comparator was chosen i.e. the standard of care for patients with chronic hepatitis C and PNALT. Indeed, the current Belgian clinical guidelines advise not to treat these patients (15).

Fourthly, guidelines advised to assess QALYs from the Belgian population. However, in our study, utility values coming from foreign people, i.e. German and American people. In the future, it would be interesting to assess them for the Belgian population. In a first stage, the description of health states could be performed using the EQ-5D generic description system. Then, health state descriptions could be

valued from the general Belgian population using the Flemish index values, estimated from a study in Flanders (19;20). Such an analysis would also allow us to use a better distribution during the probabilistic sensitivity analysis than triangular distributions for the utility values. The impact of the utilities' choice was mostly important in the first model. Indeed, confidence intervals of the ICERs in this model were extremely large, mostly because of the range of utilities tested. The question of an analysis of the value of such information is thus raised. However, to assess Belgian utilities as specified above is costly. On the other hand, the probability to take a bad decision based on the utilities used in the model is low. Indeed, the probabilistic sensitivity analysis showed that for genotypes 1-4-5-6, the ICER would be higher than €50,000 in only 1.9% of the cases (0% for genotypes 2-3) and higher than €20,000 in only 16.1% of the cases (0% for genotypes 2-3). The interest of performing such investigation to obtain Belgian utilities is thus limited.

Fifthly, we chose to extrapolate results for 35 years, which was clinically justified. However, it would have been interesting to extrapolate results for a lifelong period, as advised in the guidelines (13). Nevertheless, it should be noted that with a lifetime period, results and conclusion would have been similar.

Finally, even if we tested the elasticity of the parameters to highlight which parameters had the biggest impact on the ICER, a Tornado diagram would have been interesting to present in both models. Indeed, such a diagram ranks the model's parameters according to the magnitude of their impact on the ICER and would thus have allowed external readers of the study to immediately identify sensible parameters, even if no additional information compared to the calculation of the parameters' elasticity would be given.

References

(1) Messori A, Trippoli S, Vaiani M. Efficacy, safety, and cost of new anticancer drugs. Price needs to be evaluated against effectiveness. BMJ 2002 Nov 30;325(7375):1302.

(2) Simoons ML. Cholesterol-lowering therapy, a recommendation from the Health Council. Ned Tijdschr Geneeskd 2000 Dec 16;144(51):2442-4.

(3) Hammerschmidt T, Goertz A, Wagenpfeil S, Neiss A, Wutzler P, Banz K. Validation of health economic models: the example of EVITA. Value Health 2003 Sep;6(5):551-9.

(4) McCabe C, Dixon S. Testing the validity of cost-effectiveness models. Pharmacoeconomics 2000 May;17(5):501-13.

(5) Bennett WG, Inoue Y, Beck JR, Wong JB, Pauker SG, Davis GL. Estimates of the cost-effectiveness of a single course of interferon-alpha 2b in patients with histologically mild chronic hepatitis C. Ann Intern Med 1997 Nov 15;127(10):855-65.

(6) Sullivan SD, Jensen DM, Bernstein DE, Hassanein TI, Foster GR, Lee SS, et al. Cost-effectiveness of combination peginterferon alpha-2a and ribavirin compared with interferon alpha-2b and ribavirin in patients with chronic hepatitis C. Am J Gastroenterol 2004 Aug;99(8):1490-6.

(7) Wong JB, Koff RS. Watchful waiting with periodic liver biopsy versus immediate empirical therapy for histologically mild chronic hepatitis C. A cost-effectiveness analysis. Ann Intern Med 2000 Nov 7;133(9):665-75.

(8) Younossi ZM, Singer ME, McHutchison JG, Shermock KM. Cost effectiveness of interferon alpha2b combined with ribavirin for the treatment of chronic hepatitis C. Hepatology 1999;30(5):1318-24.

(9) Grieve S. Assessing the risk to patients of hepatitis B infection. Nurs Times 2007 Jun 26;103(26):25.

(10) Hornberger J, Farci P, Prati D, Zeuzem S, Green J, Patel KK. The economics of treating chronic hepatitis C patients with peginterferon alpha-2a (40 kDa) plus ribavirin presenting with persistently normal aminotransferase. J Viral Hepat 2006 Jun;13(6):377-86.

(11) Weinstein MC, Toy EL, Sandberg EA, Neumann PJ, Evans JS, Kuntz KM, et al. Modeling for health care and other policy decisions: uses, roles, and validity. Value Health 2001 Sep;4(5):348-61.

(12) Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. Value Health 2003 Jan;6(1):9-17.

(13) Cleemput I, Van Wilder P, Vrijens F, Huybrechts M, Ramaekers D. Recommandations pour les évaluations pharmacoéconomiques en Belgique. Health technology assessment (HTA). Bruxelles: Centre fédéral d'expertise des soins de santé (KCE); 8 A.D. Apr. Report No.: KCE Reports vol. 78B. Ref. D/2008/10.273/24.

(14) Annemans L, Warie H, Nechelput M, Peraux B. A health economic model to assess the long term effects and cost-effectiveness of PEG IFN alpha-2a in hepatitis C virus infected patients. Acta Gastroenterol Belg 2004 Jan;67(1):1-8.

(15) Michielsen P, Brenard R, Bourgeois N, de GC, Delwaide J, Henrion J, et al. Hepatitis C: screening, treatment and prevention practical guidelines. Acta Gastroenterol Belg 2003 Jan;66(1):15-9.

(16) Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales FL, Jr., et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N Engl J Med 2002 Sep 26;347(13):975-82.

(17) Hadziyannis SJ, Sette H, Jr., Morgan TR, Balan V, Diago M, Marcellin P, et al. Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. Ann Intern Med 2004;140(5):346-55.

(18) Zeuzem S, Diago M, Gane E, Reddy KR, Pockros P, Prati D, et al. Peginterferon alfa-2a (40 kilodaltons) and ribavirin in patients with chronic hepatitis C and normal aminotransferase levels. Gastroenterology 2004 Dec;127(6):1724-32.

(19) Cleemput I, Crott R, Vrijens F, Huybrechts M, Van Wilder P, Ramaekers D. Recommandations provisoires pour les évaluations pharmacoéconomiques en Belgique. Bruxelles: Centre fédéral d'expertise des soins de santé (KCE); 6 A.D. May. Report No.: KCE Reports vol. 28B. Ref. D/2006/10.273/11.

(20) Cleemput I. Economic evaluation in renal transplantation : A social preference valuation set for health states in Flanders, Belgium. Leuven: Katholieke Universiteit Leuven; 2003.

PART 3 : THE USE OF QUALITY ASSESSMENT INSTRUMENTS

1 METHODOLOGICAL CHOICES AND RESEARCH QUESTIONS

In a previous study (1), we performed a systematic review on the surgical treatment of obesity. We thus decided to test different quality assessment instruments on this review.

In the literature, various instruments are available. A previous study assessed which instruments are usually used and identified that on the 39 reviews studied, 5 reviews used the BMJ 35-items checklist, 5 used the Drummond *et al* 10-item check-list, 2 used US panel recommendations, 26 used ad hoc instruments where the number of items varied, and one review used a checklist with no clear origin (2). Moreover, according to this study, the BMJ check-list seemed to be the most complete checklist (2). We thus choose to analyze this instrument (3) and to compare it with two news instruments: the CHEC list (4;5), based on an international consensus and the QHES instrument (6), a grading system where items are weighted according to their relative importance.

Based on the experience of clinical studies, the following research questions were investigated:

- Are the instruments assessing the quality of economic evaluations appropriate for multiple designs?
- If economic evaluations were classified using a global quality score, would the classification differ according to the instrument used?
- Is the use of a global quantitative quality score needed and what is the impact of weighting the criteria according to their relative importance?

A list of all abbreviations used in this part can be found in table 1.

Table 1 : List of abbreviations

BMJ	British medical journal
CHEC list	Consensus health economic criteria list
ICC	Intraclass correlation coefficient
NA	Not appropriate
QHES	Quality of health economic studies

References

(1) Lambert ML, Kohn L, Vinck I, Cleemput I, Vlayen J, Leys M, et al. Pharmacological and surgical treatment of obesity. Residential care for severely obese children in Belgium. Brussels: Belgian Health Care Knowledge Centre (KCE); 2006 Aug. Report No.: 36C.

(2) Jefferson T, Demicheli V, Vale L. Quality of systematic reviews of economic evaluations in health care. JAMA 2002 Jun 5;287(21):2809-12.

(3) Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. BMJ 1996 Aug 3;313(7052):275-83.

(4) Ament A, Evers S, Goossens M, De Vet H, Van Tulder M. Criteria list for conducting systematic reviews based on economic evaluation studies - the CHEC project. In: Donaldson C, Mugford M, Vale L, editors. Evidence-based Health Economics. From effectiveness to efficiency in systematic review.London: BMJ Books; 2002. p. 99-113.

(5) Evers S, Goossens M, de VH, van TM, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. Int J Technol Assess Health Care 2005;21(2):240-5.

(6) Chiou CF, Hay JW, Wallace JF, Bloom BS, Neumann PJ, Sullivan SD, et al. Development and validation of a grading system for the quality of cost-effectiveness studies. Med Care 2003 Jan;41(1):32-44.

2 COMPARISON OF THREE INSTRUMENTS ASSESSING THE QUALITY OF ECONOMIC EVALUATIONS: A PRACTICAL EXERCISE ON ECONOMIC EVALUATIONS OF THE SURGICAL TREATMENT OF OBESITY

S Gerkens, R Crott, I Cleemput, JP Thissen, MC Closon, Y Horsmans, C Beguin. Comparison of three instruments assessing the quality of economic evaluations: a practical exercise on economic evaluation of the surgical treatment of obesity. International Journal of Technology Assessment in Health Care 2008;24(3):318-25.

NB: The definitive version is available at http://journals.cambridge.org/
© 2008 Cambridge University Press.

International Journal of Technology Assessment in Health Care, **24:3** (2008), 318–325.
Copyright © 2008 Cambridge University Press. Printed in the U.S.A.
doi:10.1017/S0266462308080422

Comparison of three instruments assessing the quality of economic evaluations: A practical exercise on economic evaluations of the surgical treatment of obesity

Sophie Gerkens
Université Catholique de Louvain

Ralph Crott
Cliniques Universitaires Saint-Luc

Irina Cleemput
Belgian Health Care Knowledge Centre (KCE)

Jean-Paul Thissen
Université Catholique de Louvain and *Cliniques Universitaires Saint-Luc*

Marie-Christine Closon
Université Catholique de Louvain

Yves Horsmans
Université Catholique de Louvain and *Cliniques Universitaires Saint-Luc*

Claire Beguin
Cliniques Universitaires Saint-Luc

Objectives: The increasing use of full economic evaluations has led to the development of various instruments to assess their quality. The purpose of this study was to compare the frequently used *British Medical Journal* (BMJ) check-list and two new instruments: the Consensus Health Economic Criteria (CHEC) list and the Quality of Health Economic Studies (QHES) instrument. The analysis was based on a practical exercise on economic evaluations of the surgical treatment of obesity.
Methods: The quality of nine selected studies was assessed independently by two health economists. To compare instruments, the Spearman rank correlation coefficient was calculated for each assessor. Moreover, the test–retest reliability for each instrument was assessed with the intraclass correlation coefficient (ICC) (3,1). Finally, the inter-rater agreement for each instrument was estimated at two levels: comparison of the total score of each article by the ICC(2,1) and comparison of results per item by kappa values.

We thank Maria-Isabel Portet (Researcher, Université catholique de Louvain, School of Public Health, Brussels, Belgium) for useful comments. Financial support: no funding, no sponsors. Logistic support: Université catholique de Louvain, Cliniques universitaires Saint-Luc.

Results: The Spearman's rank correlation coefficient between instruments was usually high (rho $>$ 0.70). Furthermore, test–retest reliability was good for every instruments, that is, 0.98 (95 percent CI, 0.86–0.99) for the BMJ check-list, 0.97 (95 percent CI, 0.73–0.98) for the CHEC list, and 0.95 (95 percent CI, 0.75–0.99) for the QHES instrument. However, inter-rater agreement was poor (kappa $<$ 0.40 for most items and ICC(2,1) $\leq$ 0.5).
Conclusions: The study shows that the results of the quality assessment of economic evaluations are not so much influenced by the instrument used but more by the assessor. Therefore, quality assessments should be performed by at least two independent experts and final scoring based on consensus.

Keywords: Quality assurance, Health care, Cost and cost analysis, Questionnaires, Review, Systematic, Technology assessment

With the scarcity of resources in health care, efficacy and safety become insufficient for a well-informed decision making on resources allocations. In the current environment, the priority becomes to reduce the costs without deteriorating the quality of care, or to improve quality of care at a reasonable cost (10). Consequently, interest in full economic evaluations, that is, studies comparing both costs and outcomes of at least two healthcare programs (10) has increased and numerous countries have now developed specific guidelines for economic evaluations. As a consequence, the use of systematic reviews of economic evaluations to summarize knowledge has been intensified and quality assessment instruments have been developed to evaluate the quality of published economic evaluations. The most frequently used instruments (13) are the Drummond et al. ten-item checklist (10) or the BMJ check-list (9), based on the Drummond check-list.

Jefferson et al. showed that too disparate quality assessment instruments were used (13), illustrating the need for a validated and internationally accepted list. To respond to this need, the Consensus Health Economic Criteria (CHEC) list has recently been developed (11).

On the other hand, neither the CHEC list nor the BMJ check-list were created with the aim of performing a simple comparison between studies in a quantifiable way. Quantitative measures of quality would allow ranking studies according to a quality score. One solution is to apply an equal weight for each item, but this strategy does not allow analysts to take into account the relative importance of each criterion. For this reason, a new instrument has been developed: the Quality of Health Economic Studies (QHES) (5), a grading system in which weightings differ according to the relative importance of each criterion.

The first objective of this study was to compare the BMJ check-list, the CHEC list, and the QHES instrument as quantitative tools to measure the quality of economic evaluations, and to examine the importance of weighting the criteria. The second objective was to assess the test–retest reliability and the inter-rater agreement for each instrument. Finally, problems associated with analyzing the quality of economic evaluations with these instruments were also determined and recommendations were made. The analysis was performed through a systematic review of economic evaluations of the surgical treatment of obesity.

METHODS

Quality Assessment Instruments Description

The BMJ Check-List. The BMJ set up a working party to develop a quality assessment check-list for use by both referees and authors. Drafts of the check-list were transmitted to health economists and journal editors and were debated at the "biannual meeting of the UK Health Economists' Study Group" in January 1996. The final check-list was based on a broad consensus and contains thirty-five items under three headings: study design, data collection, and analysis and interpretation of results (see Table 1). This check-list concentrates on full economic evaluations but could also be used for partial economic evaluations, or report and commentaries on economic evaluations. If items were not applicable to a specific study, a "not appropriate" (NA) response can be stated. The working party admitted that it is not possible to address all the points in the article and that authors can, for example, refer the reader to other published sources. More details about this check-list can be found in the literature (9).

The CHEC List. An initial item pool divided in nineteen categories was first developed by performing a literature search from Medline, Psychlit, Econlit, the Cochrane Library, and the National Health Service Economic Evaluation Databases (NHS EED). The criteria list was then created using the Delphi method. This method made use of a panel of expert on a specific topic to reach a consensus (21). In a first round, international experts were asked to give their opinion on the categories and the items selected from the literature research. Comments and the resulting list were redistributed among experts until a consensus was reached. Three rounds were sufficient to obtain the final criteria list. More details on the method used can be found in the literature (2;11).

The list contains nineteen yes-or-no questions (see Table 2). Authors recommended that, if not enough information was available in the article or in other published material to answer to a question, a "No" response should be

Table 1. The *British Medical Journal* Check-list[a]

Study design
The research question is stated
The economic importance of the research question is stated
The viewpoints of the analysis are clearly stated and justified
The rationale for choosing the alternative programmes or interventions compared is stated
The alternatives being compared are clearly described
The form of economic evaluation used is stated
The choice of form of economic evaluation is justified in relation to the questions addressed
Data collection
The sources of effectiveness estimates used are stated
Details of the design and results of effectiveness study are given (if based on a single study)
Details of the method of synthesis or meta-analysis of estimates are given (if based on an overview of a number of effectiveness studies)
The primary outcome measure(s) for the economic evaluation are clearly stated
Methods to value health states and other benefits are stated
Details of the subjects from whom evaluations were obtained are given
Productivity changes (if included) are reported separately
The relevance of productivity changes to the study question is discussed
Quantities of resources are reported separately from their unit costs
Methods for the estimation of quantities and unit costs are described
Currency and price data are recorded
Details of currency or price adjustments for inflation or currency conversion are given
Details of any model used are given
The choice of model used and the key parameters on which it is based are justified
Analysis and interpretation of results
Time horizon of costs and benefits is stated
The discount rate(s) is stated
The choice of rate(s) is justified
An explanation is given if costs or benefits are not discounted
Details of statistical tests and confidence intervals are given for stochastic data
The approach to sensitivity analysis is given
The choice of variables for sensitivity analysis is justified
The ranges over which the variables are varied are stated
Relevant alternatives are compared
Incremental analysis is reported
Major outcomes are presented in a disaggregated as well as aggregated form
The answer to the study question is given
Conclusion follow from the data reported
Conclusions are accompanied by the appropriate caveats

[a] See reference 9.

stated. A description of the items can be found on the Web (www.beoz.unimaas.nl/chec/). It should be noted that this list was not created to analyze the quality of economic evaluations based on modeling studies. However, in this study, all economic evaluations found in the literature were analyzed with the three lists, including modeling studies. Consequently, the quality score generated by the CHEC list for modeling studies has to be interpreted with caution.

The QHES List. A steering committee comprised of five experts in the field of health economics and three investigators developed a check-list for economic evaluations from a literature search using Medline, Healthstar, and the Cochrane databases. From existing guidelines and checklists, the committee selected 16 criteria with a "Yes" or "No" format (see Table 3). The selection was made by consensus. Then, weights for each criterion were estimated using a general linear regression (random effects) based on data collected from a conjoint analysis survey on 120 international health economists. More details about the QHES list can be found in the literature (5).

Studies Selection

The different quality assessment instruments were applied to nine economic evaluations of surgical treatment techniques of obesity. More details about the studies and the selection criteria for the studies are described elsewhere (14). While initially only full economic evaluations were selected for review, we also tested the quality assessment instruments on cost-outcome descriptions, that is, studies describing both costs and effects but not presenting an incremental cost-effectiveness ratio (ICER). Five full economic evaluations, of which two included a cost-utility and cost-effectiveness analysis and three included only a cost-utility analysis were included in the quality assessment exercise (4;7;8;19;20). Moreover, four cost-outcome description

Table 2. The Consensus Health Economic Criteria List[a]

Is the study population clearly described?
Are competing alternatives clearly described?
Is a well-defined research question posed in answerable form?
Is the economic study design appropriate to the stated objective?
Is the chosen time horizon appropriate in order to include relevant costs and consequences?
Is the actual perspective chosen appropriate?
Are all important and relevant costs for each alternative identified?
Are all costs measured appropriately in physical units?
Are costs valued appropriately?
Are all important and relevant outcomes for each alternative identified?
Are all outcomes measured appropriately?
Are outcomes valued appropriately?
Is an incremental analysis of costs and outcomes of alternatives performed?
Are all future costs and outcomes discounted appropriately?
Are all important variables, whose values are uncertain, appropriately subjected to sensitivity analysis?
Do the conclusions follow from the data reported?
Does the study discuss the generalizability of the results to other settings and patient/client groups?
Does the article indicate that there is no potential conflict of interest of study researcher(s) and funder(s)?
Are ethical and distributional issues discussed appropriately?

[a] See references 2, 11.

studies were evaluated with the three quality assessment instruments (1;6;16;17).

Quality Assessment of Economic Evaluations

The quality of selected studies was assessed independently by two heath economists (rater 1 and rater 2) using the BMJ check-list, the CHEC list, and the QHES list. Each economist blindly evaluated the quality of studies with the three instruments consecutively. Moreover, rater 1 repeated the analysis 8 weeks later. During the investigation, the guidelines of the instruments were followed and an inventory of problems associated with analyzing the quality of economic evaluations according to these guidelines was made.

Quality scores were then evaluated for each study. In a first stage, a score with an equal weight for each item was

Table 3. The Quality of Health Economic Studies Instrument[a]

Item	Weightings
Was the study objective presented in a clear, specific, and measurable manner?	7
Were the perspective of the analysis (societal, third-party payer, and so on) and reasons for its selection stated?	4
Were variable estimates used in the analysis from the best available source (i.e., Randomized Control Trial-Best, Expert Opinion-Worst)?	8
If estimates came from a subgroup analysis, were the groups prespecified at the beginning of the study?	1
Was uncertainty handled by: (i) statistical analysis to address random events; (ii) sensitivity analysis to cover a range of assumptions?	9
Was incremental analysis performed between alternatives for resources and costs?	6
Was the methodology for data abstraction (including value health states and other benefits) stated?	5
Did the analytic horizon allow time for all relevant and important outcomes? Were benefits and costs that went beyond 1 year discounted (3–5%) and justification given for the discount rate?	7
Was the measurement of costs appropriate and the methodology for the estimation of quantities and unit costs clearly described?	8
Were the primary outcome measure(s) for the economic evaluation clearly stated and were the major short term, long term and negative outcomes included?	6
Were the health outcomes measures/scales valid and reliable? If previously tested valid and reliable measures were not available, was justification given for the measures/scales used?	7
Were the economic model (including structure), study methods and analysis, and the components of the numerator and denominator displayed in a clear transparent manner?	8
Were the choice of economic model, main assumptions and limitations of the study stated and justified?	7
Did the author(s) explicitly discuss direction and magnitude of potential biases?	6
Were the conclusions/recommendations of the study justified and based on the study results?	8
Was there a statement disclosing the source of funding for the study?	3

[a] See reference 5.

calculated as a quantitative proxy for the evaluation's quality. In a second stage, the importance of weighting the criteria was examined. For the QHES instrument, weights determined by Chiou et al. were used (5). For the BMJ and the CHEC list, no weighting exist. Implicit weightings determined by one of the assessor according to the relative importance he confers to each item (subjective assessment) were thus used. In summary, three types of quality scores were obtained: a score without weighting of the criteria, a score with an implicit weighting for the BMJ and the CHEC list, and a score with an explicit weighting determined by Chiou et al. for the QHES instrument.

Statistical Analysis

To compare the instruments, the range and the mean of the quality scores generated by each instrument were calculated. Ranking differences between instruments were then assessed using the Spearman rank correlation coefficient. In this study, we considered a correlation coefficient of $r \geq 0.7$ as high, $0.7 > r \geq 0.5$ as moderate, and $r < 0.5$ as low.

Moreover, test–retest reliability between time 1 and 2 was assessed for each instrument by the rater 1 using model 3 of the six ICCs discussed by Shrout and Fleiss (18), that is, the ICC(3,1) where raters are assumed to be representative of the entire population of raters (Supplementary Figure 1, available at www.journals.cambridge.org/thc).

Finally, for each instrument, the inter-rater agreement at time 1 was estimated at two levels: comparison of the total score of each article by the ICC(2,1) (18), where raters are assumed to be a random subset of all possible raters, and comparison of results per item by kappa values. Kappa values less than 0.40, between 0.40 and 0.74, and between 0.75 and 1 were defined as poor, fair to good, and perfect agreement respectively (12;15). Tests were performed using SAS software version 9.

RESULTS

Instruments Comparison

The comparison of instruments showed that they mainly analyze similar items. Nevertheless, some differences can be highlighted (see Table 4).

First, only the BMJ check-list investigates if the economic importance of the study question is stated and if the choice of alternatives is justified. On the other hand, the BMJ check-list does not assess if the choice of cost and outcome items is appropriate, as it is done in the other two instruments. Finally, this is the only instrument that does not include a question about the presence of conflicts of interest by the authors.

Second, the CHEC list is designed for clinical trial and observational studies. Consequently, there is no item on model characteristics. Moreover, this instrument does not determine whether limitations of the studies are specified.

Table 4. Instruments' Comparisons

	BMJ	CHEC	QHES
Objective	Y	Y	Y
Economic importance of the study question	Y	N	N
Economic study design	Y	Y	N
Description of the population	Y	Y	N
Subgroup analysis	N	N	Y
Perspective	Y	Y	Y
Time horizon and discount rate	Y	Y	Y
Alternative description	Y	Y	N
Alternative choice	Y	N	N
Outcomes choice	N	Y	Y
Outcomes measurement	Y	Y	Y
Outcomes valuation	Y	Y	Y
Costs choice	N	Y	Y
Costs measurement	Y	Y	Y
Costs valuation	Y	Y	N
Details of the model	Y	N	Y
Incremental analysis	Y	Y	Y
Handle of uncertainty	Y	Y	Y
Results presentation	Y	N	N
Appropriateness of the conclusion	Y	Y	Y
Limitations	Y	N	Y
Results generalizability	N	Y	N
Ethical aspect	N	Y	N
Conflict of interest	N	Y	Y

BMJ, British Medical Journal; CHEC, Consensus Health Economic Criteria; QHES, Quality of Health Economic Studies; Y, handled; N, not handled.

On the other hand, it is the only instrument asking whether ethical aspect and generalizability of the results are investigated.

Third, the QHES instrument determines if subgroups analyzed are appropriately defined but does not examine whether details on the population and on the study design are specified, in contrast to both the BMJ check-list and the CHEC list. Finally, this instrument does not investigate if details about price adjustments for inflation or about currency conversion are given.

Inventory of Problems Encountered During the Quality Assessment

It was often difficult to judge from the studies if an item was respected or not because too little information was given in the publication. For example, details on cost calculations were regularly limited and sometimes, only sources were given. In such situation, it is thus important to consult these sources to be able to evaluate the quality of the studies with more accuracy.

Moreover, the BMJ check-list and the QHES instruments were mainly adapted to modeling studies while the CHEC list was conceived for clinical trials and observational studies. Consequently, the item assessing if details of the model were given was, for example, not adapted to clinical trials and

observational studies. Hence, it could be interesting to have an instrument adapted to several study designs with specific subquestions for each design.

It was also often difficult to choose between a "Yes" or "No" response. Some items regrouped various criteria. Consequently, if only one of the criterion was not respected, a "no" response should be stated, even if other criteria were respected. The possibility to use an intermediate value as "partially respected" could thus be interesting. This problem was mostly present with the QHES instrument. For example, one item tested if the time horizon was relevant, if costs and outcomes were discounted and if the discount rate was justified. It would be interesting to test the impact of subdividing this kind of item.

Differences in Quality Scores Between Instruments

With equal weights between items, the quality score of the nine selected studies and the three ratings (rater 1, time 1; rater 1, time 2; rater 2, time 1) varied between 30.8 and 90.0 of 100 points on the BJM check-list, between 15.8 and 84.2 of 100 points on the CHEC list, and between 6.3 and 87.5 of 100 points on the QHES instrument. Means and standard deviations of the studies quality scores for each instrument and the three ratings are detailed on the Web site (Supplementary Table 5, available at www.journals.cambridge.org/thc).

With a weighting between items, the quality score of the nine selected studies and the three ratings varied between 6 and 92 of 100 points on the BJM check-list, between 14 and 89 of 100 points on the CHEC list, and between 22 and 77 of 100 points on the QHES instrument.

Hierarchical ranking of studies

The hierarchical ranking of studies based on their quality score and the Spearman ranking correlation coefficient can be found at the Web site (Supplementary Tables 6 and 7, available at www.journals.cambridge.org/thc). A high Spearman ranking correlation between instruments was found, except for rater 1 in time 1 where the correlation was moderate between the BMJ and the CHEC list.

Moreover, for each instrument, the Spearman ranking correlation coefficients between weighted and not weighted scores were high and ranged from 0.83 to 0.99. Thus, weighting of criteria has little impact on the ranking. On the other hand, the ranking varied according to the assessor, as shown by the inter-rater agreement.

Test–retest Reliability and Inter-rater Agreement

Test–retest reliability in terms of ICC(3,1) was good for all instruments, that is, 0.99 (95 percent CI, 0.86–0.99) for the BMJ check-list, 0.97 (95 percent CI, 0.73–0.98) for the CHEC list, and 0.95 (95 percent CI, 0.75–0.99) for the QHES instrument. However, there was poor inter-rater agreement. For the BMJ check-list, agreement was poor for twenty-seven of thirty-five items (77 percent), fair to good for six of thirty-five items (17 percent), and perfect for only two out of thirty-five items (6 percent). For the CHEC list, agreement was poor for fifteen of nineteen items (79 percent) and perfect for four of nineteen items (21 percent). For the QHES instrument, agreement was poor for ten of sixteen items (63 percent), fair to good for three of sixteen items (19 percent), and perfect for three of sixteen items (19 percent). Overall inter-rater agreement in terms of ICC(2,1) was 0.52 (95 percent CI, 0.21–0.83) for the BMJ check-list, 0.33 (95 percent CI, 0.07–0.71) for the CHEC list, and 0.33 (95 percent CI, 0.02–0.73) for the QHES instrument.

DISCUSSION

Instruments comparisons highlighted the subjective character of the quality assessment. Indeed, results were not influenced by the instruments used but rather by experts who analyzed the studies. As shown in the study, instruments mainly assessed similar items, which could explain the high Spearman rank correlation coefficient.

The poor agreement between experts could be explained by various factors. First, time spent to analyze studies might have an impact on results. One author spent around 1 day per study to assess deeply the quality of the study and returned systematically to referred sources if insufficient details were provided in the basic article. The other expert spent around 2 days to assess the quality of all the studies and based his analysis on the main article only.

Second, the subjectivity of the examinants could also influence the response. A complete respect of criteria was rare and intermediate responses were not authorized. Severe raters could have tendency to state a 0 value if one criterion of the item was not completed, while another rater could have tendency to state a 1 value, considering that on the whole, the criteria were respected.

Third, experience of the rater in economic evaluation could also play a role. One rater has worked in the health economics domain for nearly 20 years, while the experience of the second rater was only 2 years. Thus, it is possible that they consider the quality of the studies differently.

Fourth, the perception and interpretation of the items and the ambiguity of the responses also influenced the results. Items were sometimes large and could be interpreted in various ways. Some items also referred to specific study design and when the design of the study was not appropriate, reaction of raters could differ.

It should also be noted that the BMJ check-list and the CHEC list were created as qualitative instruments and not as scoring instruments. On the other hand, calculating a quality score from these instruments allowed us to easily have an idea of the ranking of these studies according to their quality.

Finally, caution is needed when interpreting our results given that the limited number of studies led to high confidence intervals for the inter-rater agreement. In a previous study, two people analyzed the quality of 30 economic evaluations of health promotion with the QHES instrument and found a better inter-rater reliability (IC95 percent: 0.64–0.91) (3). However, our results plead in favor of doing further research to estimate the overall role of the evaluator in assessing the quality of economic evaluations. To do so, an international study including a higher number of evaluators and in particular a larger sample of studies from various areas should be conducted.

In conclusion, our findings highlight that in practice, results are not so much influenced by the instrument used but more by the assessor. It is thus essential to perform quality analyses of economic evaluations by at least two blinded experts and to base the final scoring on a consensus. Moreover, a clear definition of each item should be given and respected by raters. Experts should also spend a substantial period of time to analyze studies thoroughly and should refer to sources of information when specified in the article if not enough details are provided in the basic study. Finally, in the future, it would be interesting to create a single instrument adapted to each study design and to introduce the possibility to use an intermediate score value.

CONTACT INFORMATION

Sophie Gerkens, MSc, PhD Candidate (sophie.gerkens@uclouvain.be), Health Economist, School of Public Health–Unité de Socioéconomie de la santé (SESA), Université Catholique de Louvain, 30 Clos Chapelle-aux-Champs box 3041, Brussels 1200, Belgium
Ralph Crott, PhD (ralph.crott@uclouvain.be), Health Economist, Department of Medicine, Cliniques Universitaires Saint-Luc, 10, Av. Hippocrate, Brussels B-1200, Belgium
Irina Cleemput, PhD (Irina.Cleemput@kce.fgov.be), Expert Economic Analysis, Department of Research, Belgian Health Care Knowledge Centre (KCE), 62 Rue de la Loi, Brussels B-1040, Belgium
Jean-Paul Thissen, MD, PhD (Thissen@diab.ucl.ac.be), Professor, Department of Endocrinology, Université Catholique de Louvain, 10, av. Hippocrate, Brussels B-1200, Belgium; Chief, Department of Endocrinology, Cliniques Universitaires Saint-Luc, 10, av. Hippocrate, Brussels B-1200, Belgium
Marie-Christine Closon, PhD (closon@sesa.ucl.ac.be), Professor, School of Public Health – Unité de socioéconomie de la santé (SESA), Université Catholique de Louvain, 30 Clos Chapelle-aux-Champs, Box 3041, Brussels B-1200, Belgium
Yves Horsmans, MD, PhD (yves.horsmans@uclouvain.be), Professor, Department of Gastroenterology, Université Catholique de Louvain; Chief, Department of Gastroenterology, Cliniques Universitaires Saint-Luc, 10, Av. Hippocrate, Brussels B-1200, Belgium
Claire Beguin, MD, PhD (claire.beguin@uclouvain.be[), Chief, Department of Medical Information and Statistics, Cliniques Universitaires Saint-Luc, 10, Av. Hippocrate, Brussels B-1200, Belgium

REFERENCES

1. Agren G, Narbro K, Jonsson E, et al. Cost of in-patient care over 7 years among surgically and conventionally treated obese patients. *Obes Res*. 2002;10:1276-1283.
2. Ament A, Evers S, Goossens M, De Vet H, Van Tulder M. Criteria list for conducting systematic reviews based on economic evaluation studies – the CHEC project. In: Donaldson C, Mugford M, Vale L, eds. *Evidence-based health economics. From effectiveness to efficiency in systematic review*. London: BMJ Books; 2002:99-113.
3. Au F, Prahardhi S, Shiell A. Reliability of two instruments for critical assessment of economic evaluations. *Value Health*. 2007. In press.
4. Chevallier JM, Daoud F, Szwarcensztein K, Volcot MF, Rupprecht MF. Medicoeconomic evaluation of the treatment of morbid obesity by Swedish adjustable gastric banding (SAGB). *Ann Chir*. 2006;131:12-21.
5. Chiou CF, Hay JW, Wallace JF, et al. Development and validation of a grading system for the quality of cost-effectiveness studies. *Med Care*. 2003;41:32-44.
6. Christou NV, Sampalis JS, Liberman M, et al. Surgery decreases long-term mortality, morbidity, and health care use in morbidly obese patients. *Ann Surg*. 2004;240:416-423.
7. Clegg AJ, Colquitt J, Sidhu MK, et al. The clinical effectiveness and cost-effectiveness of surgery for people with morbid obesity: A systematic review and economic evaluation. *Health Technol Assess*. 2002;6:1-153.
8. Craig BM, Tseng DS. Cost-effectiveness of gastric bypass for severe obesity. *Am J Med*. 2002;113:491-498.
9. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ*. 1996;313:275-283.
10. Drummond MF, Sculpher MJ, Torrance JW, O'Brien BJ, Stoddart JL. *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press; 2005.
11. Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on health economic criteria. *Int J Technol Assess Health Care*. 2005;21:240-245.
12. Fleiss JL. *Statistical methods for rates and proportions*. New York: John Wiley and Sons; 1981.
13. Jefferson T, Demicheli V, Vale L. Quality of systematic reviews of economic evaluations in health care. *JAMA*. 2002;287:2809-2812.
14. Lambert ML, Kohn L, Vinck I, et al. *Pharmacological and surgical treatment of obesity. Residential care for severely obese children in Belgium*. Brussels: Belgian Health Care Knowledge Centre (KCE); 2006. Report 36C.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.

16. Martin LF, Tan TL, Horn JR, et al. Comparison of the costs associated with medical and surgical treatment of obesity. *Surgery*. 1995;118:599-606.
17. Nguyen NT, Goldman C, Rosenquist CJ, et al. Laparoscopic versus open gastric bypass: A randomized study of outcomes, quality of life, and costs. *Ann Surg*. 2001;234:279-289.
18. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 1979;86:420-428.
19. van Gemert WG, Adang EM, Kop M, et al. A prospective cost-effectiveness analysis of vertical banded gastroplasty for the treatment of morbid obesity. *Obes Surg*. 1999;9:484-491.
20. van Mastrigt GA, van Dielen FM, Severens JL, Voss GB, Greve JW. One-year cost-effectiveness of surgical treatment of morbid obesity: Vertical banded gastroplasty versus Lap-Band. *Obes Surg*. 2006;16:75-84.
21. Whitman NI. The Delphi technique as an alternative for committee meetings. *J Nurs Educ*. 1990;29:377-379.

Additional tables available on the web site

http://journals.cambridge.org/
© 2008 Cambridge University Press.

Table 5 Means and standard deviations of the studies quality scores for each instrument

Rater / Time	**BMJ**		**CHEC**		**QHES**	
	No weighting	Weightings	No weighting	Weightings	No weighting	Weightings
Rater 1 / Time 1	51.4 +/- 21.3	43.9 +/- 20.4	39.8 +/- 21.7	40.6 +/- 23.2	38.2 +/- 23.9	39.9 +/- 24.4
Rater 1 / Time 2	56.4 +/- 20.8	48.1 +/- 20.1	45.0 +/- 20.9	47.0 +/- 21.9	45.1 +/- 22.3	45.0 +/- 24.1
Rater 2 / Time 1	63.9 +/- 10.5	58.4 +/- 11.9	65.5 +/- 16.2	68.2 +/- 19.1	54.2 +/- 10.8	57.1 +/- 10.5

Table 6 Hierarchical ranking of studies

Rater / Time	References	BMJ		CHEC		QHES	
		No weighting (Ranking)	Weightings (Ranking)	No weighting (Ranking)	Weightings (Ranking)	No weighting (Ranking)	Weightings (Ranking)
Rater 1 / Time 1	[13]	86.67 (1)	70.00 (2)	78.95 (1)	83.00 (1)	81.25 (1)	84.00 (1)
	[10]	83.87 (2)	77.00 (1)	57.89 (2)	55.00 (3)	56.25 (3)	59.00 (2)
	[11]	61.29 (3)	60.00 (3)	52.63 (3)	56.00 (2)	56.25 (2)	56.00 (3)
	[12]	50.00 (4)	44.00 (4)	42.11 (5)	44.00 (5)	43.75 (4)	46.00 (4)
	[14]	39.29 (5)	35.00 (5)	15.79 (8)	15.00 (8)	25.00 (6)	24.00 (7)
	[9]	38.71 (6)	32.00 (6)	26.32 (6)	28.00 (6)	6.25 (9)	6.00 (9)
	[15]	37.04 (7)	32.00 (7)	15.79 (9)	14.00 (9)	25.00 (7)	28.00 (6)
	[16]	34.62 (8)	23.00 (8)	21.05 (7)	20.00 (7)	12.50 (8)	15.00 (8)
	[17]	30.77 (9)	22.00 (9)	47.37 (4)	50.00 (4)	37.50 (5)	41.00 (5)
Rater 1 / Time 2	[13]	90.00 (1)	73.00 (2)	84.21 (1)	89.00 (1)	87.50 (1)	92.00 (1)
	[10]	83.87 (2)	76.00 (1)	57.89 (3)	59.00 (2)	68.75 (2)	71.00 (2)
	[11]	67.74 (3)	68.00 (3)	52.63 (4)	56.00 (4)	43.75 (4)	43.00 (4)
	[12]	62.50 (4)	53.00 (4)	57.89 (2)	58.00 (3)	50.00 (3)	51.00 (3)
	[14]	48.15 (5)	40.00 (5)	26.32 (7)	28.00 (7)	37.50 (6)	32.00 (7)
	[9]	45.16 (6)	37.00 (6)	36.84 (6)	40.00 (6)	12.50 (9)	14.00 (9)
	[15]	40.74 (7)	36.00 (7)	21.05 (9)	23.00 (8)	37.50 (7)	38.00 (6)
	[16]	38.46 (8)	28.00 (8)	21.05 (8)	20.00 (9)	25.00 (8)	22.00 (8)
	[17]	30.77 (9)	22.00 (9)	47.37 (5)	50.00 (5)	43.75 (5)	42.00 (5)
Rater 2 / Time 1	[13]	81.25 (1)	74.00 (2)	78.95 (4)	83.00 (4)	62.50 (2)	63.00 (2)
	[11]	75.00 (2)	74.00 (1)	84.21 (1)	88.00 (1)	75.00 (1)	76.00 (1)
	[9]	67.74 (3)	61.00 (4)	78.95 (2)	85.00 (3)	56.25 (3)	58.00 (5)
	[10]	65.63 (4)	65.00 (3)	78.95 (3)	86.00 (2)	56.25 (5)	62.00 (3)
	[12]	63.64 (5)	59.00 (5)	68.42 (5)	75.00 (5)	56.25 (4)	61.00 (4)
	[14]	60.00 (6)	54.00 (6)	57.89 (6)	57.00 (6)	43.75 (8)	44.00 (8)
	[17]	60.00 (7)	54.00 (7)	52.63 (8)	50.00 (8)	50.00 (7)	53.00 (7)
	[15]	56.67 (8)	48.00 (8)	52.63 (7)	53.00 (7)	50.00 (6)	56.00 (6)
	[16]	45.16 (9)	37.00 (9)	36.84 (9)	37.00 (9)	37.50 (9)	41.00 (9)

Table 7 Spearman rank correlation coefficient

Weightings	Instruments	Rater 1 / Time 1	Rater 1 / Time 2	Rater 2 / Time 1
No weighting	BMJ / CHEC	0.65	0.78	0.94
	BMJ / QHES	0.76	0.71	0.93
	CHEC / QHES	0.82	0.88	0.91
Weightings	BMJ / CHEC	0.59	0.78	0.90
	BMJ / QHES	0.76	0.72	0.92
	CHEC / QHES	0.80	0.88	0.87

3 CONCLUSION AND PERSPECTIVES

The major finding in this part concerned the fact that classification of economic evaluations according to their quality scores were more influenced by the assessor rather than by the instrument used, showing the importance of analyzing the quality of economic evaluations by at least two researchers and to base results on a consensus. Compared to instruments evaluating the quality of clinical trials, we thus found different results. The classification of economic evaluations according to their quality did not differ significantly between the three studied instruments, which respond to our second research question.

Moreover, an analysis of the possible reasons for disagreement between raters allowed us to determine other methodological improvement possibilities.

Firstly, the disagreement between raters could be due to differences in their experience, showing the necessity of having at least two experts. This raised the following question: Which background and how many years of experience should be required to correctly assess the quality of economic evaluations? It should be interesting to deal with this point in future researches.

Secondly, the time spent to analyze the data could also explain the disagreement between the assessors. Consequently, it is important that at least one of the assessors analyzes the economic evaluations thoroughly. Moreover, poor reporting was an important problem raised during our analysis. So, if not enough information is available in the study, at least one of the assessors has to return to the

publicized sources of data mentioned in the study and if possible, to contact the authors of the study in order to obtain the original data.

Then, because a complete respect of the items' criteria was rare, the subjectivity of the assessors may also play a role. Indeed, they could only choose between a "yes" or "no" response for the CHEC list and the QHES instrument, and between a "yes", "no" or "not appropriate" response for the BMJ check-list. Consequently, when the item was partially respected, one assessor could consider that the item was not respected while the other could consider that on the whole, it was respected. To reduce the impact of the subjectivity, it would thus be interesting to assess the possibility to give intermediary responses, e.g. to respond that the item was only partially respected and to give only half of the points.

Moreover, because some items were not adapted to the study design, reactions of the assessors could differ. A solution could be either to create one instrument with specific sub-questions adapted to the study design, or else to use different instruments according to the study design. In addition, because the interpretation of the items may differ, a clear description of the item should be given and respected. Such a point also respond to our first research questions, i.e. the fact that instruments were not adapted to each study design. However, it should be noted that this had no impact on the classification of the studies.

The impact of weighting the items according to their relative importance was also studied and no classification difference was found between weighted items and un-weighted items. So, in our study, use of a weighting of the item has no significant impact on the results. However, it should be noted that for the BMJ check-list and the CHEC list, implicit weightings determined by one of the assessors

were used and the impact of such choices on the studies' classification was not tested. In the future, it would be interesting to conduct a conjoint analysis with international health economists in order to estimate weightings for the BMJ check-list and the CHEC list from general linear regression analyses, as in the study of Chiou *et al* (1). As information, table 1 shows items where a consensus between instruments was found and the related weightings determined by Chiou *et al* (1).

Tables 1. Consensus between instruments and related weightings (determined by Chiou *et al* (1))

Items	Weightings (QHES) (1)
Objective	7
Perspective	4
Time horizon and discount rate	7
Outcomes measurement	6
Outcomes valuation	7
Costs measurement	8
Incremental analysis	6
Handle of uncertainty	9
Appropriateness of the conclusion	8

On the other hand, the question of using a global quantitative quality score remains. Such a score allows one to classify the studies according to their quality. However, the number of full economic

evaluations identified in a systematic review is usually limited and different populations and interventions are assessed. Consequently, even if different conclusions were found between the studies, not enough information would be available to determine if results differed according to the quality of the studies or according to the study design, which is also a problem in meta-analyses. Calculating a global quantitative quality score instead of a narrative quality assessment is thus less needed if few economic evaluations have been identified and if similar conclusions are found between the studies. But with the increasing interest in economic evaluations, the necessity of calculating a global quantitative quality score could increase. Moreover, even if few studies are identified in the systematic review, the assessment of their quality, at least in a narrative way, remains important. Indeed, if all studies are of poor quality, it could be hazardous to base a decision on them.

Moreover, even if discussions about a global quality score remain, these instruments could be used by authors in the publication process to check if their study contains enough data to allow other researchers to correctly assess the quality of the study. During our analysis, assessors highlighted the problem of poor reporting in economic evaluations. Poor reporting had already been noticed in previous studies and increases the difficulty of researchers to assess the quality of economic evaluations. This problem could be reduced by the use of quality assessment instruments before publication. On the other hand, it is not always possible to provide all data in a study, given journals' restrictions in the number of pages. A first solution would be to include some points in separate studies and to give the reference towards these studies. Assessors would thus have to refer to these additional studies during the quality assessment process. Another solution would be to require, for every economic evaluation,

a standard technical report as supplementary information which would be available on internet.

Finally, our study had some limitations. Indeed, only two assessors investigated the quality of a limited number of economic evaluations. In the future, a larger international study with more assessors and more economic evaluations from various disease areas could be conducted in order to confirm our results. Moreover, the impact of a blinding of the assessors for the authors and the journal of the studies would be interesting to assess, such a situation not having been tested in our analysis. The impact of the background and the experience years of the assessors could also be interesting to test.

References

(1) Chiou CF, Hay JW, Wallace JF, Bloom BS, Neumann PJ, Sullivan SD, et al. Development and validation of a grading system for the quality of cost-effectiveness studies. Med Care 2003 Jan;41(1):32-44.

DISCUSSION

Improving of information provided by economic evaluations is sometimes difficult to perform due to the costs generated by such improvement. Indeed, designs such as Practical Clinical Trials would increase the external validity of results compared to economic evaluations performed through RCTs, but they are often too expensive to conduct, especially for the analysis of events with a low frequency of occurrence or for chronic diseases. Usually, a trade-off between the quality of the information and its cost must be made. In this thesis, three relatively inexpensive tools that overcome the limitations of RCTs were investigated: the use of administrative databases, the use of Markov models and the use of quality assessment instruments. Both clinical and methodological recommendations as well as perspectives for future research could be highlighted by such investigations.

Firstly, the use of routinely collected Belgian administrative databases allowed us to measure effectiveness criteria in term of incidences of events when assessing prophylaxis treatments in prevention of VTE after major orthopedic surgery. In the future, it could be interesting to use such data to compare hospitals' practices and to establish benchmarks, as well as to compare prophylaxis treatments by estimating incremental cost-effectiveness ratio. On the other hand, important data as the patient’s weight and non-reimbursed pharmacological treatments could not be identified from these databases and other sources of data such as the patients’ records

had to be investigated. These investigations allowed us to determine that in the analyzed hospital, improving of the adherence to guidelines was needed, showing the interest of implementing a continuous quality improvement cycle and of creating computer-based alert systems to detect patients at risk. The fact that other sources of data had to be investigated also revealed that methodological improvements were needed. With an improving link with other databases such as data from the patients' medical records and on outpatients' drug consumptions, the reliability of observational studies from routinely collected databases would increase and quality of care criteria such as the adherence to guidelines could be measured. The inclusion of non reimbursed pharmacological treatments in the current databases should also be encouraged.

In a second stage, modelling studies were used to combine both effectiveness and efficacy data and to determine from which conditions a treatment effective in RCTs would become ineffective in the daily practice. The evaluation of a treatment with Markov models is mostly useful for chronic illnesses. Consequently, we decided to focus on chronic hepatitis C disease.

Our study has shown that the treatment of patients with moderate chronic hepatitis C and PNALT should be promoted. On the other hand, the interest to immediately treat genotypes 1-4-5-6 patients with mild chronic hepatitis C should be confirmed in the future. Moreover, the predictive value of these models should be validated by conducting a prospective observational study for these patients. Then, from a methodological point of view, with the development of routinely collected databases, the access to real world data needed to construct Markov models would also be improved.

Finally, the use of instruments to assess the quality of care was analyzed and showed that they permit to improve the information available to health care decision makers at two levels. Firstly, by using them in the beginning of a peer review process, they improve the quality of economic evaluations in terms of method used and reporting. Secondly, by using them during a systematic review, they should allow researchers to identify the strengths and weaknesses of the studies, to measure if the methods used are appropriated, to determine if the results are pertinent and to assess if there are related to the setting of the review.

Numerous instruments could be used in the literature. However, our study has shown that the quality evaluation was more influenced by the assessor rather than by the instrument used, highlighting the importance of assessing the quality of economic evaluations by at least two health economists. Moreover, at least one of the experts should assess the studies quality in a detailed way by systematically referring to the sources mentioned and if needed, by contacting the authors in case of poor reporting. The interest of demanding a technical report which would be available on internet should also be discussed. Concerning the use of a global quality score, even if such a score would allow a simple comparison of the studies, we showed that is not necessary to weight the items according to their relative importance. However, our results should be confirmed by conducting an international research with more raters and more studies from various areas. The interest of developing one single instrument adapted to every study design, to allow for intermediary responses, and to blind the assessors should also be tested in the future. Finally, the impact of the rater's experience and background would be

interesting to assess in order to determinate a minimum number of years of experience and a minimum background required.

In conclusion, this thesis allowed us to suggest methodological recommendations in the field of health economics but also clinical recommendations. Improvements of the information available to health care decision makers could thus be done but there is still a long way to go in this area. In the future, it is important to promote the following points:

Firstly, information provided by economic evaluations is mostly used in decision processes concerning the reimbursement of health care interventions. However, this information could also be utilized to determine clinical practice guidelines. For instance, the information provided by Markov models developed in this thesis could help in the determination of new clinical guidelines concerning the treatment of chronic hepatitis C.

Secondly, information provided by Belgian databases could be exploited to improve the health care quality. Indeed, in Belgium, the health care system especially focuses on budget and reimbursement policies but less on the assessment of the quality, which is a major drawback (1). A rapport of the OCDE assessing a number of health care quality indicators showed that among the 28 member countries, Belgium was placed between the 15th and the 20th place for each indicator (1;2). So, efforts concerning the assessment and the improvement of the health care quality should be consented.

The health care quality can be measured by structural indicators assessing the inputs of the system (i.e. the qualification of the physicians, the hospitals' equipments, etc) but also by indicators based on processes (i.e. indicators regarding if health care are delivered in an appropriate way), and by indicators based on results (i.e. indicators that compare health care results such as mortality rates) (1). Such indicators should be implemented at a national and even international level. Currently, some Belgian hospitals and health care institutions begin to assess quality indicators individually but a national plan should be coordinated. For instance, the national council for the promotion of the quality (Conseil National de la Promotion de la Qualité (CNPQ)) currently focuses on recommendations on good clinical practices and gives statistics in terms of overconsumption. This council could also set up a systematic quality assessment process in which indicators on both processes and results would be compared at a national and even at an international level, and could publish the comparisons for both health care institutions and citizens. Our thesis has shown that Belgian hospital databases could be used to implement such kind of assessment concerning the incidence of VTE after major orthopedic procedures. In the future, it would be interesting to determine if such kind of evaluation could also be applied for other diseases and to implement them systematically.

Thirdly, a development of registries and an improved linkage between existing Belgian health care databases should be supported. Indeed, Real World data are essential in economic evaluations and our thesis has shown that important efforts in this sector should be consented. Moreover, with such improvements, the assessment of quality

indicators based on processes and on results could be performed more easily, which is expected to increase the Belgian health care quality.

Fourthly, the participation in international working research groups and the follow-up of international recommendations concerning new health economic methods should be encouraged. For instance, it would be interesting to submit our study comparing quality assessment instruments to the International Society For Pharmacoeconomics and Outcomes Research (ISPOR) in order to create an international research groups on this topics and to set up an internationally accepted instrument, which will help in the standardization of health economic methods and guidelines at an international level.

Fifthly, a multidisciplinary approach and an improved dialogue between health care experts (physicians, health economists,…) should be promoted. For instance, health economists must take into account the experience of the physicians and the problems they face in their practices when conducting economic evaluations. Conversely, it is also important that physicians consider the results of economic evaluations in their clinical practice.

Finally, it is also important to address the fact that economic evaluations are necessary but insufficient sources of information and that other clinical, political and ethical aspects must be considered.

References

(1) Daue F, Crainich D. Quel est l'état de santé du système de santé en Belgique? Brussels: Itinera Institute; 2008 Apr 29.

(2) De Vos M. Le modèle social belge n'est plus un modèle. Brussels: Itinera Institute; 2008 May 2.

NOTES

All the parts of this thesis have been published:

- S Gerkens, C Beguin, R Crott, MC Closon, Y Horsmans. Assessing the quality of pharmacological treatments from administrative databases: The case of low-molecular-weight heparin after major orthopedic surgery. Journal of Evaluation in Clinical Practice 2008 Aug;14(4): 585-94.

- S Gerkens, R Crott, I Cleemput, JP Thissen, MC Closon, Y Horsmans, C Beguin. Comparison of three instruments assessing the quality of economic evaluations: a practical exercise on economic evaluation of the surgical treatment of obesity. International Journal of Technology Assessment in Health Care 2008;24(3):318-25.

- S Gerkens, M Nechelput, L Annemans, B Peraux, M Mouchart, C Beguin, Y Horsmans. A health economic model to assess the cost-effectiveness of PEG IFN α-2a and ribavirin in patients with mild chronic hepatitis C. Journal of viral hepatitis 2007 Aug;14(8):523-36.

- S Gerkens, M Nechelput, L Annemans, B Peraux, C Beguin, Y Horsmans. A health economic model to assess the cost-effectiveness of pegylated interferon α-2a and ribavirin in patients with moderate chronic hepatitis C and persistently normal alanine aminotransferase levels. Acta gastroenterological belgica 2007 Apr-Jun;70(2):177-87.

All the parts of this thesis have also been presented at international conferences:

- S Gerkens, C Beguin, MC Closon, I Cleemput, JP Thissen. Cost-effectiveness studies of the surgical treatment of obesity : the need for methodological improvement. EUPHA 15th European Conference on Public Health. October 11-13th 2007. Helsinki, Finland.

- S Gerkens, R Crott, C Beguin, MC Closon, Y Horsmans. Comparison of three instruments assessing the quality of economic evaluations. ISPOR 10th annual European Congress. October 20-23 th 2007. Dublin, Ireland.

- S Gerkens, C Beguin, R Crott, MC Closon,Y Horsmans. Assessing the quality of care from medical basic datasets: The case of low-molecular-weight heparin after major orthopedic surgery. 23 rd PCS International Conference. November 7-10 th 2007. Venice, Italy.

- S Gerkens, M Nechelput, L Annemans, B Peraux, C Beguin, Y Horsmans. A health economic model to assess the cost-effectiveness of pegylated interferon α-2a and ribavirin in patients with moderate chronic hepatitis C and persistently normal alanine aminotransferase levels. XIXth Belgian Week of Gastroenterology. 08-10 février 2007. Ostend, Belgium [Poster].

List of abbreviations

ALT	Alanine transaminase
APR-DRG	All patient refined diagnosis related group
BMJ	British medical journal
CHEC list	Consensus health economic criteria list
CI	Confidence interval
CNK	Code national(e) kode
DRG	Diagnosis related group
DVT	Deep-vein thrombosis
EVR	Early viral response
FPR	Fibrosis progression rate
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HFS	Hip fracture surgery
HIV	Human immunodeficiency virus
ICC	Intraclass correlation coefficient
ICD-9-CM	International classification of diseases 9th revision, clinical modification
ICER	Incremental cost-effectiveness ratio
LMWH	Low-molecular-weight heparin
LOS	Length of stay
MBDS	Minimum basic data set
MDC	Major diagnosis category
NA	Not appropriate
NSAIDs	Non-steroidal anti-inflammatory drugs
OR	Odds ratio
PCR	Polymerase chain reaction
PE	Pulmonary embolism
PEG IFN	Pegylated interferon
PNALT	Persistently normal alanine transaminase
QALY	Quality-adjusted life-year
QHES	Quality of health economic studies
RIZIV/INAMI	National health institute (Rijksinstituut voor ziekte en invaliditeitsverzekering /Institut national d'assurance maladie invalidité)
RNA	Ribonucleic acid
SVR	Sustained viral response
SVT	Thrombosis of superficial vein
THR	Total hip replacement
TKR	Total knee replacement
VAT	Value added tax
VTE	Venous thromboembolism