UNIVERZA V LJUBLJANI

FAKULTETA ZA FARMACIJO

BARBARA DEBELJAK

DIPLOMSKA NALOGA

RAZVOJ IN UPORABA METOD ZA OCENJEVANJE PRENOSLJIVOSTI FARMAKOEKONOMSKIH ŠTUDIJ

DEVELOPMENT AND USAGE OF METHODS FOR ASSESSING TRANSFERABILITY OF PHARMACOECONOMIC STUDIES

UNIVERZITETNI ŠTUDIJ FARMACIJE

Ljubljana, 2011

Univerza
v Ljubljani
Fakulteta
za farmacijo

BARBARA DEBELJAK

DIPLOMSKA NALOGA

RAZVOJ IN UPORABA METOD ZA OCENJEVANJE PRENOSLJIVOSTI FARMAKOEKONOMSKIH ŠTUDIJ

DEVELOPMENT AND USAGE OF METHODS FOR ASSESSING TRANSFERABILITY OF PHARMACOECONOMIC STUDIES

UNIVERZITETNI ŠTUDIJ FARMACIJE

Ljubljana, 2011

Diplomsko nalogo sem opravljala na Fakulteti za farmacijo, na Katedri za biofarmacijo in farmakokinetiko pod mentorstvom prof. dr. Aleša Mrharja, mag. farm v okviru sodelovanja s Faculty of Health, Medicine and Life Sciences, Maastricht University, Netherlands pod vodstvom somentoric: izredne profesorice dr. Silvie Evers in dr. Saskie Knies, Department of Health Organization, Policy and Economics.

Ker je raziskovalno delo potekalo v angleškem jeziku in v sodelovanju s tujegovorečimi strokovnjaki, bo diplomska naloga v celoti napisana v angleškem jeziku.

IZJAVA

Izjavljam, da sem diplomsko nalogo samostojno izdelala pod vodstvom mentorja prof. dr. Aleša Mrharja, mag. farm. ter somentoric dr. Silvie Evers and dr. Saskie Knies.

Barbara Debeljak

Ljubljana, 2011

Predsednik komisije: prof. dr. Marija Sollner Dolenc, mag. farm.

Član komisije: doc. dr. Mitja Kos, mag. farm.

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

ABSTRACT

Many jurisdictions now request economic evaluations as part of their decision-making procedures for the pricing and reimbursement of pharmaceuticals and other health technologies. Decision makers may not have economic evaluations with local data inputs available, and may wish to use studies already performed in other settings. Furthermore performing all pharmaco-economic evaluations locally is not always efficient. In the last decades several methods have been developed for assessing the transferability of economic evaluations. Data, methods and results are transferable if a potential user can assess their applicability to their setting and they are applicable to that setting. The aim of this thesis is to describe issues around transferability of health economic studies, and to test one of the recently published methods for assessing transferability. For the purpose of testing the method, whose concept was published by Boulenger (2005) and supplemented in detail by Nixon (2009), an economic evaluation of (bivalent or quadrivalent) HPV vaccination of girls only in pre-sexual period alongside cervical cancer screening versus screening alone. The method evaluates the level of reporting of the items most relevant for transferability, but does not assess what the ICER value would be in another setting. A systematic literature search of economic evaluations of HPV vaccination was first performed. The transferability of a total of thirty economic evaluations was assessed. Studies had in general moderate quality of reporting items relevant for transferability. Furthermore, the interpretation and scoring process were considered as subjective, and some items were not only difficult to interpret but we also had difficulties to apply them on specific case of HPV vaccination. The lowest level of reporting was assessed for effectiveness and benefit measures, while detailed information about discounting and study perspective were better provided. Basic costs data like currency, included costs, price year etc. were well reported. On the other hand details about sources of cost values were assessed as insufficient. Based on our research work we could see that the checklist is not well suited for model-based studies, as only two items of the checklist applied on the model properties and that further specification of items in this section would be useful. In future, research of factors influencing transferability of economic evaluations should be performed, especially their relative importance should be investigated. Guidelines for transparent and comprehensive

reporting of economic evaluations should be developed, specifically tailored for model-based and trial-based economic evaluations.

POVZETEK

UVOD

V zadnjem obdobju se v vse več državah pri odločanju o financiranju zdravil zahteva predložitev farmakoekonomskih študij. Farmakoeknomske študije z lokalnimi podatki niso vselej na voljo, zato je v takih primerih prisotna težnja po uporabi študij izvedenih v drugih državah. Poleg tega izvedba farmakoekonomskih študij na lokalni ravni ni vedno učinkovita, ne samo iz finančnih razlogov, ampak tudi zaradi pomanjkanja podatkov, časa in človeških virov. Podatki, metode in rezultati študije so prenosljivi, kadar posameznik lahko oceni njihovo aplikativnost na svoje okolje in lahko le-te tja tudi prenese. Številni faktorji, kot so na primer značilnosti zdravstvenega sistema, stroški, epidemiologija bolezni, razpoložljivost virov v zdravstvu, različna klinična praksa, demografske značilnosti itd. vplivajo na prenosljivost farmakoekonomskih evalvacij iz enega okolja v drugo. Nedavno tega je bila v številnih državah aktualna odločitev o umestitvi cepljenja proti humanem papiloma virusu v zdravstveni sistem. Humani papiloma virus je povzročitelj predrakavih in rakavih sprememb na materničnem vratu. Na svetu vsako leto zboli za rakom materničnega vratu 470.000 in umre 233.000 žensk. Obstaja več serotipov humanega papiloma virusa. Trenutno sta na voljo dve cepivi, ki preprečujeta okužbo s serotipoma 16 in 18, ki povzročata 70% vseh primerov raka materničnega vratu. Uvedba omenjenega cepljenja v klinično prakso predstavlja velik finančni zalogaj za zdravstvene sisteme po vsem svetu, zato so odločitve o uvedbi cepljenja temeljile tudi na njegovi ekonomski upravičenosti oz. stroškovni učinkovitosti. Znatne razlike v epidemiologiji, klinični praksi in stroških zdravljenja so povečale potrebe po lokalnih farmakoekonomskih podatkih in postavile v ospredje vprašanje prenosljivosti že obstoječih farmakoekomskih študij.

NAMEN

Namen diplomske naloge bo predstaviti področje prenosljivosti farmakoekonomskih študij in aplikacija ene izmed objavljenih metod za vrednotenje prenosljivosti (t.i. točkovnik

EURONHEED) na primeru študij stroškovne učinkovitosti cepljenja proti humanem papiloma virusu.

METODE

V zadnjem desetletju so bile objavljene metode oziroma poskusi vrednotenja prenosljivosti farmakoekonomskih študij. V diplomski nalogi smo na praktičen primer aplicirali tako imenovani točkovnik EURONHEED, ki je bil prvič objavljen leta 2005 s strani Boulenger in sod. ter dopolnjen leta 2009 s strani Nixon in sod. Omenjeni točkovnik z dvainštiridesetimi vprašanji ocenjuje prenosljivost metodologije in podatkov, uporabljenih v farmakoekonomski študiji, na način, da preverja kvaliteto in doslednost poročanja ključnih informacij. Avtorji točkovnika so tudi določili šestnajst vprašanj, ki so jih ocenili kot najbolj pomembne pri ocenjevanju prenosljivosti med posameznimi okolji. Vsako vprašanje se točkuje z vrednostmi 0, 0,5 ali 1, oziroma se označi kot neuporabno. Na koncu se izračuna delež možnih točk. Za izbrani študijski primer smo na sistematičen način v podatkovnih bazah MEDLINE in NHS EED poiskali objavljene farmakoekonomske študije, jih selekcionirali na podlagi predhodno določenih vključitvenih kriterijev ter jih točkovali.

REZULTATI

Na podlagi iskalnega profila smo identificirali 159 zadetkov, izmed katerih smo glede na vključitvene kriterije v točkovanje vključili 30 študij. V povprečju so imele študije ocenjeno kakovost poročanja na srednjem nivoju. Razpon doseženih točk (delež vseh možnih točk) je segal od 23,3% do 79,0%, povprečna vrednost je znašala 61,7%. V primeru dela točkovnika, ki vsebuje 16 najbolj pomembnih vprašanj o prenosljivosti, je bilo skupno število doseženih točk višje. Povprečna vrednost je znašala 75,3%, razpon pa je bil od 41,7% do 100%.

Nekatera vprašanja smo s težavami interpretirali in točkovali. Aplikacija metode ocenjevanja prenosljivosti se je izkazala kot relativno subjektiven način evaluiranja, poleg tega so se določena vprašanja slabo nanašala na naš izbrani primer. Najmanj izčrpen nivo

poročanja je bil ocenjen za podatke o učinkovitosti in mere koristnosti, medtem ko so bili podatki o diskontni stopnji in vidiku raziskave bolje opisani. Osnovni podatki o stroških, kot so valuta, vključeni stroški, leto obračunanih stroškov itd. so bili ocenjeni kot dobro poročani, medtem ko so bili viri stroškov slabše poročani.

ZAKLJUČKI

Na podlagi našega primera je razvidno, da je točkovnik EURONHEED manj ustrezen za modelne farmakoekonomske študije. Le dve izmed 42-ih vprašanj se namreč nanašata na informacije o modelu. Nadaljna specifikacija zlasti tega dela vprašalnika bi bila dobrodošla.

Izračun doseženih točk predstavlja relativno grobo oceno prenosljivosti. Točkovanje je subjektivne narave, še posebno pri vprašanjih, ki niso dovolj specifično definirana. Ravno tako je interpretacija končnega rezultata težavna, saj avtorji metode ne navajajo kaj posamezna vrednost točk v praksi pomeni (pri kateri vrednosti se študija smatra kot prenosljiva).

Potrebno je nadaljno raziskovanje faktorjev, ki vplivajo na prenosljivost, zlasti njihova relativna pomembnost. Ključnega pomena je tudi nadaljni razvoj smernic, ki bi določale transparentno in izčrpno poročanje v objavljenih farmakoekonomskih študijah, ter zajemale vse vidike prenosljivosti. Smernice bi bilo smiselno oblikovati ločeno za modelne študije in za farmakoekonomske raziskave, ki se izvajajo vzporedno s kliničnimi.

LIST OF ABBREVATIONS

HTA = Health technology assessment

EUnetHTA = European Network of Health Technology Assessment

INAHTA = International Network of Agencies for Health Technology Assessment

ISPOR = International Society of Pharmacoeconomics and Outcomes Research

QALY = Quality adjusted life year

ICER = Incremental cost-effectiveness ratio

EURONHEED = European Network of Health Economic Evaluation Databases

NHS EED = National Health Service Economic Evaluation Database

MEDLINE = Medical Literature Analysis and Retrieval System Online

HPV = Human papillomavirus

CIN = Cervical Intraepithelial Neoplasia

ICC = Invasive cervical carcinoma

FIGO = the International Federation of Gynecology and Obstetrics

PAP smear test = Papanicolaou smear test

IMF = International Monetary Foundation

1. INTRODUCTION

Healthcare systems have developed at different speeds and with differing degrees of complexity throughout the previous decades, reflecting the diverse political and social conditions in each country. Despite their diversity all systems share a common reason for their existence, namely the improvement of health for their entire populations. To attain this goal a healthcare system undertakes a series of functions, most notably the financing and delivering of healthcare services. Since available resources are limited, delivering health services involves making decisions. Decisions are required on what interventions should be offered, the way the health system is organized, and how the interventions should be provided in order to achieve an optimal health gain with available resources, while, at the same time, respecting people's expectations. Decision makers thus need information about the available options and their potential consequences. (1)

Over the last decade, Health Technology Assessment (HTA) has developed as a field of scientific research to inform policy and clinical decision making around the introduction and diffusion of health technologies. (2, 3)

HTA is a multidisciplinary field that systematically evaluates the effects of a technology on health, on the availability and distribution of resources, and on other aspects of healthcare system performance, such as equity and responsiveness. HTA addresses economic, organizational, social, and ethical impacts of a new technology. (1) Health technology is considered as a broad term that encompasses medicines, medical devices, diagnostic tests, medical and surgical procedures, and other clinical, public health and organizational interventions. (2, 3)

The beginning of HTA in Europe can be dated back to the late 1970s, when interest in the economic aspects of health technologies started to grow, and the first scientific activities in the evaluation of health interventions in terms of HTA can be identified. (4) Since the beginning of HTA activities, efforts have been made at international level to share experiences. Today, networks exist at the European level (European Network of Health

Technology Assessment (EUnetHTA)), as well as outside Europe (International Network of Agencies for Health Technology Assessment (INAHTA)). (2, 3)

1.1. Transferability of economic evaluations

A health economic analysis is an integral part of an HTA assessment. Two or more alternative treatment strategies are compared in terms of their costs and benefits, and the result is expressed as an incremental cost-effectiveness ratio (ICER). ICER represents additional costs that are produced to obtain an additional unit of health. New interventions with a high ICER value can be denied reimbursement on the grounds of not providing enough benefit for the additional resources needed to treat a patient with a new intervention.

HTA agencies throughout Europe now request from manufacturers to submit an economic evaluation when applying for drug reimbursement. As more and more jurisdictions request economic studies, the burden on health technology manufacturers and researchers increases, particularly when the various national guidelines insist on the presentation of local data, or the use of specific methods. (5)

However, performing studies locally is not always feasible. This limitation is particularly typical for small countries and for low and middle income countries with limited resources for carrying out economic evaluations. Consequently, decision makers in these countries may not have economic evaluations with local data inputs available, and may wish to use studies already performed in other settings. This is time and cost saving compared to conducting a completely new study for their local environment. (6, 7)

Decision-makers can use economic evaluations from other jurisdictions in two ways: a) by applying the conclusions directly because the results are either assumed or assessed to be relevant for their local setting; b) by using the methods and data that are applicable and substituting local methods and data for those that are not. (8) Namely, studies may be considered ***generalizable*** if they can be applied to a range of jurisdictions without any

adjustment needed for interpretation and ***transferable*** if they can be adapted to apply them to other settings. (10)

The published literature reveals that transferability has no explicit and unambiguous definition, and it is often not clearly distinguished from generalizability. (8, 9) Some published concepts of transferability issues refer directly to **results** of the economic evaluations (study), others to **data** used in economic evaluations and some even to **methods** performed in economic evaluations. Particularly transferability of methods was addressed in a recently published paper. (10)

However, some of the authors write about the transferability of the studies in a very general manner, and it might be usually understood that results of the study are addressed. (11) For example, Späth et. al. (1999) consider that the **results** of a study are **transferable** if potential users can assess whether the results apply to their settings and adapt them if necessary. (6) Welte et. al. (2004) do not explicitly define concept of **transferability**, although they implicitly refer to it as the capacity to use results of the economic evaluations obtained in study country to another (decision) country. (9, 12) Drummond et. al. (2005) consider economic evaluations to be **generalizable** if the **results** of an HTA undertaken in one country are relevant to another. In addition to this statement they state that studies performed in other setting cannot be applied without adaptation in another location, and are most of use in the setting where they were performed. (13) Boulenger et al. (2005) define **generalizability** as the degree to which **results** of an economic analysis hold true in other settings, while **transferability** is defined as: **data**, **methods and results** are transferable, if a potential user can assess their applicability to their setting and they are applicable to that setting. They understand transferability as a broader concept comparing to generalizability. (8) Mason and Mason (2006) concept of **generalizability** encompass three elements*:* ***technical merit*** (authors emphasize the meaning of transparent reporting, use of best methods, best-quality evidence etc.), ***applicability*** (to policy context) and ***transferability,*** where the latter denotes capacity to directly use the findings in other settings in a reliable way. (14) Goeree et al. (2007) interpret **transferability** in terms of using **data** of economic evaluations from one geographic area to another. Transferability is interpreted as a very geographic concept and is not precisely distinguished to

generalizability. Several terms such as transferability, generalizability, portability and extrapolation have been used to describe when data from an economic evaluation done in one geographic area is transferred to another location or transferred across time. (15) According to the ISPOR (International Society of Pharmacoeconomics and Outcomes Research) Task Force's working definitions are economic evaluations **generalizable** if they apply, without adjustment, to other settings. **Data** are **transferable** if they can be adapted to apply to other settings. (10) Barbieri et al. (2010) gives a very similar definition. Namely, studies may be considered as **generalizable** if they can be applied to a range of jurisdictions without any adjustment needed for interpretation. **Studies** are **transferable** if they can be adapted to apply to other settings. (11) The definition published by Boulenger et. al. (8) seems to be the most explicitly defined and also covers the widest range of transferability issues concerning data, results and methods used in economic evaluations. Furthermore, it is in line with recently published ISPOR Good Research Practice Task Force's report. (10) Therefore, use of the term *transferability* in the thesis is based on the definition by Boulenger et. al.: **transferability** is defined as: **data**, **methods and results** are transferable, if a potential user can assess their applicability to their setting and they are applicable to that setting. Also, the method (EURONHEED transferability checklist, described below) used in our case to assess transferability of studies applies to the same definition.

1.2. Factors limiting transferability of economic evaluations

In the literature several reasons for limiting transferability are mentioned, which might vary from location to location. Main factors affecting transferability are shown in Figure 1.

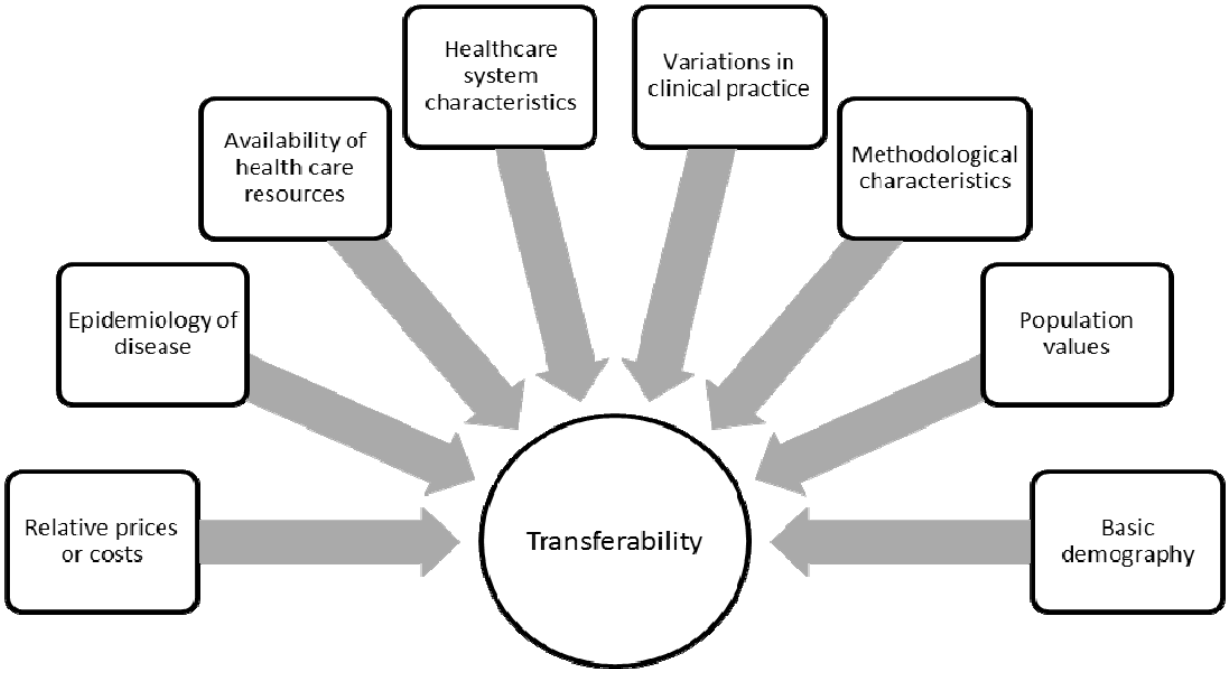

Figure 1: Factors affecting transferability (adapted from Goeree (15), Drummond (17))

Goeree et al. (2007) summarized and categorized the literature on factors affecting transferability of economic evaluation data. They developed a classification system which grouped 77 factors into five broad categories based on the characteristics of the patient, disease, provider, health care system and methodological interventions. Details about this classification system are presented in Table I.

Table I: Classification of factors potentially affecting transferability (15)

Patient characteristics	Disease characteristics	Provider characteristics	Health care system characteristics	Methodological characteristics
Demographics (age, gender, race), education, socio-economic status	Epidemiology (incidence/prevalence, disease progression, spread)	Clinical practice, conventions, guidelines, norms	Absolute or relative prices	Costing methodology, estimation procedures
Risk factors, medical history, genetic factors	Disease severity, case mix	Experience, education, training, skills, learning curve position	Available resources (staff, facilities, equipment), programs, services	Study perspective
Lifestyle, genetic factors, environmental factors	Disease interaction, co-morbidity, concurrent medications	Quality of care provided	Organization of delivery system, structure, level of competition	Study factors (artificial trial conditions, industry-related bias)
Attitudes toward treatment, culture, religion, hygiene, nutrition	Mortality due to disease	Method of remuneration (supplier-induced demand)	Level of technological advancement, innovation and availability	Timing of the economic evaluation
Life expectancy, mortality rates		Patient identification	Available treatment options (comparators)	Clinical outpoints/outcome measures
Compliance and adherence rates, ethical standards		Cultural attitudes	Capacity utilization, economies of scale, technical efficiency	Discount rates
Income, employment rates, productivity, work loss time, friction time		Incentives for providers, liability	Access to programs and services	Exchange rates, purchasing power parties
Population density, immigration and emigration patterns			Market form of suppliers, payment of suppliers, suppliers incentives	Opportunity cost (foregone benefits)
Population values			Waiting lists, referral patterns	Affordability (CE thresholds)
Type of insurance coverage, user fees, co-payments			Regulatory and organizational infrastructure, licensing of products	
Incentive of patients			Availability of generics or substitutes	
			Specialization of labour	
			Incentives for institutions	

1.3. Development of methods for assessing transferability

Methods to assess transferability and variability in economic evaluation studies have been discussed extensively in the literature relating to both trial-based and modelling studies. (6) In the last decades several methods have been developed for assessing the transferability of economic evaluations.

One of the first general methods developed to decide which studies are transferable is the **model of Welte**. Welte's model is a decision chart method that includes general and specific knock-out criteria to assess whether the study is transferable, and how the ICER would change in another setting. (12)

Knock-out criteria can be described as criteria that make the transfer of study results always impossible or so troublesome that conducting a new study is a better option. (12) Thus, by using general and specific knock-out criteria it can be determined which studies can be transferred to the decision country and which not. Detailed scheme of the decision chart is presented in the Appendix. Three **general knock-out criteria** are: 1) the evaluated technology is not comparable to the one that shall be used in decision country, 2) the comparator is not comparable to the one that is relevant in the decision country (for example a comparator drug is not licensed in the decision country) and 3) the study does not possess an acceptable quality. If any of the general knock-out criteria apply then it is impossible to transfer study results from one country to another. If the study passes these general knock-out criteria, the specific knock-out criteria are used to determine which parts of the studies are transferable. Specific knock-out criteria were defined with the help of transferability factors identified through literature review, and were systematically grouped. Details about potential transferability factors included in Welte's model are presented in Table II.

Factors or **specific knock-out criteria** in the transferability checklist are used to determine the correspondence between the study country and the decision country. This is carried out in three steps. First, the relevance of each transferability factor on the incremental cost-effectiveness ratio (ICER) in the study country is determined. In the second step, the level

of correspondence between the study country and the decision country for this transferability factor is estimated. At last, the expected effect of the factor on the ICER in the decision country is assessed. When e.g. the factor is relevant and the correspondence between the study country and decision country is high, the change of the ICER in the decision country will be small. In the case of a low correspondence, the estimation of the ICER will be biased, and the correct ICER might be either higher or lower, depending on the estimated direction of change.

After these three steps, it can be decided which (modelling) adjustments are necessary to transfer the foreign studies. Modelling adjustments are always necessary when there are big differences between the study country and the decision country. Namely new resource utilisation and valuation, or even modifications of the original model structure may need to be performed. Further details about the decision chart are illustrated in the Appendix.

Table II: Specific knock-out criteria of Welte's model

Group of transferability factors	Potential transferability factors
Methodological characteristics	perspective, discount rate, medical cost approach, productivity cost approach
Healthcare system characteristics	absolute and relative prices in healthcare, practice variation, technology availability
Population characteristics	disease incidence/prevalence, case-mix, life expectancy, health-status preferences, acceptance, compliance, incentives to patients, productivity and work-loss time, disease spread

Welte's model enables identification of the most needed adjustments. Its advantage is that it helps to prioritise adjustments by importance and determine what data should be primarily gathered. Authors concluded that this method also shows that the more complex the study is, the more effort is required to assess the transferability. They applied their decision chart to three case studies and predicted they might have missed some specific issues in transferability. Furthermore, they stated that their approach might be rather pragmatic than purely scientific as it includes descriptive estimation of the transferability factors.

Boulenger et. al. (2005) developed a checklist containing concepts which should be considered in the economic evaluation studies to ensure their transferability. The objective of the checklist development was to provide a tool for *assessing the level of reporting transferability information*, which is fundamental to any assessment of transferability of economic evaluations. (8)

The checklist consists of forty-two items related to overall quality of economic evaluation and helps to assess the level of reporting of transferability information. The full checklist represents the general quality checklist. The forty-two different questions are divided into six main sections: subject and key elements of the study (questions Q_1-Q_2, HT_1-HT_2, SE_1-SE_2,P_1, SP_1-SP_4, M_1-M_2), characteristics of the methods measuring clinical outcomes (questions E_1-E_7), measure of the health benefits used in the economic analysis (questions B_1-B_5), costs (questions C_1-C_{11}), discounting (questions D_1-D_4), and discussion by the authors (questions S_1, O_1). The checklist is provided in Table III.

Table III: Transferability information checklist provided by Boulenger et al. (8)

	Transferability information checklist question
Q1	Is the study question clearly stated?
Q2	Are the alternative technologies justified by the author(s)?
HT1[a]	Is the intervention described in sufficient detail?
HT2[a]	Is (are) the comparator(s) described in sufficient details?
SE1	Did the authors correctly specify the setting in which the study took place (e.g. primary care, community)?
SE2[a]	Is (are) the country state which the economic study took place clearly specified?
P1[a]	Did the authors correctly state which perspective they adopted for the economic analysis?
SP1[a]	Is the target population of the health technology clearly stated by the authors or when it is not done can it be inferred by reading the article?
SP2	Are the population characteristics described? (e.g. age, sex, health status, socio-economic status, inclusion/exclusion criteria)
SP3[a]	Does the article provide sufficient detail about the study sample(s)?
SP4	Does the paper provide sufficient information to assess the representativeness of the study sample with the respect to the target population?
M1	If a model is used is it described in detail?
M2	Are the origins of the parameters used in the model given?
E1	If a single study is used is the study design described (sample selection, study design, allocation, follow-up)?
E2	If a single study is used are the methods of data analysis described (ITT/per protocol or observational data)?

E3	If based on a review/synthesis of previous published studies, are review methods described (search strategy, inclusion criteria, sources, judgment criteria, combination, investigation of differences)?
E4	If based on opinion, are the methods used to derive estimates described?
E5[a]	Have the principal estimates of effectiveness measures been reported?
E6	Are the side effects or adverse effects addressed in the analysis?
E7[a]	Does the article provide the results of a statistical analysis of the effectiveness results?
B1	Do the authors specify any summary benefit measure(s) used in the economic analysis?
B2	Do the authors report the basic method of valuation of health states or interventions?
B3	Do the authors specify the source(s) of health states (e.g. Specific patient population or the general public)?
B4	Do the authors specify the valuation tool used?
B5[a]	Is the level of reporting of benefit data adequate (incremental analysis, statistical analyses)?
C1[a]	Are the cost components/items used in the economic analysis presented?
C2	Are the methods used to measure costs components/items provided?
C3	Are the sources of resource consumption data provided?
C4	Are the sources of unit price data provided?
C5[a]	Are unit prices for resources given?
C6[a]	Are costs and quantities reported separately?
C7[a]	Is the price year given?
C8	Is the time horizon given for each element of the cost analysis?
C9[a]	Is the currency unit reported?
C10	Is a currency conversion rate given?
C11	Does the article provide the results of a statistical analysis of cost results?
D1	Was the summary benefit measure(s) discounted?
D2	Were the costs data discounted?
D3	Do the authors specify the rate(s) used in discounting costs and benefits?
D4	Were discounted and not discounted results reported?
S1[a]	Are quantitative and/or descriptive analysis conducted to explore variability from place to place?
O1[a]	Did the authors discuss caveats regarding the generalizability of their results?

[a]Items comprising the transferability subchecklist.

Additionally, a subchecklist was suggested, consisting of the some of the above mentioned items which were considered to be the most important for assessing the transferability. The full checklist of 42 items represents general quality checklist, when the subchecklist of 16 items represents checklist assessing transferability properties of the study. The thought was that, whilst a study could have a high score on the overall checklist, it may be deficient in several important areas. The authors of the method independently selected a subset of questions they felt were most essential to judge the transferability of a paper. The subchecklist contains sixteen questions: HT_1, HT_2, SE_2, P_1, SP_1, SP_3, E_5, E_7, B_5, C_1, C_5, C_6,

C_7, C_9, S_1 and O_1. From the checklist with 42 questions all the items about health techonolgies, perspective, half of the items about study population and almost half of questions from cost section were selected as most important items for assessing transferability of the studies and were therefore included in the subchecklist.

For the scoring purpose, the answers to each question are classified as: *'yes'*, *'partially'*, *'no or no information provided'* and *'not applicable'*. Responses are given the score: 1 for *'yes'*, 0.5 for *'partially'* and 0 for *'no or no information provided'*. When the answer is *'not applicable'*, the question is excluded from the scoring by reducing the denominator in the summary formula. A summary score is derived using the following formula:

$$Scores\ (\%) = \frac{1}{n-x}\sum_i S_i \times 100,$$

where $i = 1,.., n$, n is the number of questions, x is the number of questions for which response was N/A and S is the score of each question.

The score reflects how thoroughly and consistently key methodological items have been addressed and how precisely they are reported. Contrary to the method of Welte, this method does not evaluate how the ICER would change in another setting.

The checklist exists of a large number of items, whereby all of them are equally important. In addition, the reporting criteria are very loosely specified, which puts the user of the checklist to uncertainty while determining the values of scores. Therefore, in 2009 Nixon et al. (20) published guidelines for completing the checklist presented by Boulenger et al (2004) and called it the EURONHEED transferability information checklist. In their paper they provided supplementary instructions so that the overall and subchecklist can be more consistently utilised.

In 2008 Antonanzas et. al. (21) published a comment on existing methods for assessing transferability, and introduced a new transferability index for published economic evaluations. They extended principles presented in studies of Boulenger et al. and Welte et

al. in a way to obtain measurement of the transferability that not only summarises the results, but also *weighs items* according to their relative importance, and considers the possibility of stopping the checking process when some critical factors are identified (as suggested by Welte et al.). Authors propose the assessment of the General Transferability Index, which represents a degree of *general transferability*, where critical and non-critical objective factors of a given study are determined and weighted. Critical objective factors are factors leading to initial exclusion of the study if the quality of the study is low or if the relevant parameters needed to calculate the cost-effectiveness ratio are not provided. Additionally, the Specific Transferability Index is defined as a degree of *specific transferability*, representing the level of difficulty that exists in applying or adapting the information from the original study to the new setting. Authors applied the method on a set of economic evaluations of infectious diseases. To gain the weights for non-critical factors they sent their questionnaire to seven different HTA agencies in Spain. They have shown that the degree of applicability of the second index depends on the characteristics of the decision maker and his ability to use the information contained in the original study. Therefore, the Specific Transferability Index must be tailored to the specific setting where the original study is to be applied. (21)

Since our research group at Maastricht University has already had experience with the method of Welte et al. (18) but not others we decided to test the EURONHEED checklist that was proposed by Boulenger et. al. and further specified by Nixon et. al. The full checklist as described by Nixon et. al. is provided in the Methods section.

An economic evaluation of Human papilloma virus (HPV) vaccination was chosen as a study case with regards to the importance of this recent health care topic in Slovenia and worldwide. Due to the high costs of vaccination, country specific epidemiology of HPV infections and cervical cancer, and benefits of vaccination occurring in the distant future, a large number of health economic evaluations was anticipated. In this respect, brief information about the HPV vaccines, cervical cancer, and models used for evaluating cost-effectiveness of HPV vaccinations is provided in the following section.

1.4. Cervical cancer and HPV vaccination

Worldwide, the incidence of cervical cancer is 470,000 new cases and 233,000 deaths per year; it is the second leading cause of cancer deaths in women, with 80% observed in developing countries. Around Europe the incidence of cervical cancer is approximately 10-15 women per 100 000 women yearly. (22)

It has been found that infection with HPV is a necessary, although not sufficient cause of cervical cancer. (23) HPV is primarily spread through sexual contact and is associated with a wide range of diseases, including cervical, vaginal, vulvar, anal, penile, head and neck cancers, as well as anogenital warts and recurrent respiratory papillomatoses. (24)

Persistent infection with cancer-associated HPV (termed oncogenic or high-risk HPV) cause the majority of squamous cell cervical cancer, the most common type of cervical cancer, and its precursor lesions, the low-grade cervical dysplasia Cervical Intraepithelial Neoplasia-1 (CIN1) and moderate-to-high-grade dysplasia (CIN2/3). Multiple HPV strains cause varying degrees of invasive cervical cancer (ICC) and its CIN precursors. HPV strains 16 and 18 cause approximately 70% of all cervical cancers and CIN3, 50% of CIN2 cases and 35 to 50% of all CIN1. Low-oncogenic HPV risk types 6 and 11 account for 90% of genital wart cases. (25)

Cervical cancer screening programs, such as the use of routine screening via the Papanicolaou (PAP) cervical smear, have substantially reduced the incidence and mortality of ICC in developed countries over the past 50 years. (25, 26) However, there has been a slowing of these declines in recent years due to poor sensitivity of cervical cytology, anxiety and morbidity of screening investigations, poor access to and attendance of screening programs, falling screening coverage, and poor predictive value of adenocarcinoma, an increasingly common cause of invasive cervical carcinoma (ICC). (26)

Recently, two different HPV vaccines were approved and are already available on the market - a bivalent vaccine (Cervarix®) which prevents infection with oncogenic types 16 and 18, and a quadrivalent vaccine (Gardasil®) which prevents also infection with types 6

and 11. Type 6 and 11 are the main cause of genital warts. Both vaccines were proven to be safe and highly effective against type-specific persistent infection. Phase 3 trials showed that the bivalent vaccine was 100% effective against HPV types 16 and 18, and the quadrivalent vaccine was 98% effective. (27, 28) Based on the proportion of CIN lesions and invasive cancer that is attributable to HPV types 16 and 18, vaccination was assumed to reduce approximately 35% of CIN 1 lesions, 51% of CIN 2/3 lesions, and 66% of ICC. (29) HPV vaccines, however, do not offer full protection against cervical cancer, as they do not protect against all oncogenic HPV strains, such as e.g. strains 31 and 45, which are also implicated in ICC and cervical dysplasia. In addition, due to the limited follow-up, long-term efficacy is still uncertain but within 5 years no reduction of efficacy was observed and HPV type-specific antibody levels remained at high level. (30)

1.5. Modelling cost-effectiveness of the HPV vaccination

There are two types of economic evaluations (31):

a) Clinical trial based economic evaluations: using patient-level data collected alongside randomized controlled trials (analysis of patient records or charts)

b) Model based economic evaluations: based on secondary analysis of data using decision analytical modelling

Decision modelling is increasingly used as a tool in economic evaluations, particularly where there is a specific resource allocation decision to be taken. The value of a formal analytic framework for decision making is that it offers a means of synthesizing available evidence from a range of sources rather than relying on a single study. It provides a way of relating the available evidence to the specific decision problem being posed. (31)

In health economic evaluation, models are typically used in two situations. First, where the relevant clinical trials have not been conducted or did not capture costs, decision analytic models are used to synthesize the best available data. Second, where the clinical trials

measure intermediate endpoints or have only short-term follow-up, models are used to extrapolate beyond the trial to final endpoints. (32)

Modelling frameworks include formalised approaches such as decision trees, Markov models, discrete event simulation and system dynamics. Models vary in complexity and resources required. The choice of approach depends on the characteristics of the disease, the impact of the technology, and the availability of data for its assessment. No single framework is always applicable. (33)

In the case of cervical cancer, simulation models can be used to translate short-term findings from vaccine trials into predictions of long-term outcomes. Two types of mathematical models have been used to evaluate the long-term effectiveness of a vaccination programme: Markov/cohort models and transmission dynamic models.

Cohort models are used to track the costs and outcomes associated with a group of individuals of identical age over time, whereby individuals transit across different health states defined in the model. Cohort models concern either a cohort of specific size (e.g., 1000 females) or unspecific size where the focus is on estimating the changing probabilities of being in different states.

In this type of model, each individual can reside in only one health state at any point in time and transitions occur from one health state to the other at defined intervals of equal length according to transition probabilities based on population characteristics (age, sexual risk, HPV type). The process can be repeated until the entire cohort has advanced to the death state and then survival time and healthcare costs are calculated with the time spent in each compartment over the lifetime of the cohort. For HPV, the different states are usually susceptible (have not been infected), infected, CIN and invasive cervical cancer (Figure 2). Some models may also include screening and treatment compartments that modify the transition probabilities.

On the other hand, a dynamic model, instead of following a single cohort, tracks a changing population over time. In a dynamic model, individuals constantly enter the model

as they are born and exit as they die, which means that, as long as people are being born, the model does not have a natural stopping point. In this type of model, individuals are susceptible, infected or immune (immunized individuals who have recovered from an infection or have been vaccinated), or progress on to CIN and invasive cervical cancer.

The main difference between Markov and dynamic modelling is that the latter accounts for the HPV vaccination reducing the prevalence of infection in the population over time. Thus, dynamic modelling has the advantage of properly assessing the impact of "herd immunity". Herd immunity means protection of the non-vaccinated individual due to a reduction in the transmission of infection (to significantly reduce the rate of cervical cancer in the population as a whole, about 70 % of girls need to be vaccinated to achieve herd immunity (36)). From the cost-effectiveness perspective, accounting for the herd immunity leads to lower values of the incremental cost-effectiveness ratio.

Another advantage of dynamic modelling is that it allows more flexibility, such as accounting for type-specific HPV and individual-based risk factors, permits the risk of future events to depend on one or more prior events, and allows evaluation of a female-only versus a male-female vaccination programme. Nevertheless, dynamic transmission models require more model parameters, which increase the level of model uncertainty. (34, 35) Figure 3 represents an example/a scheme of a dynamic model.

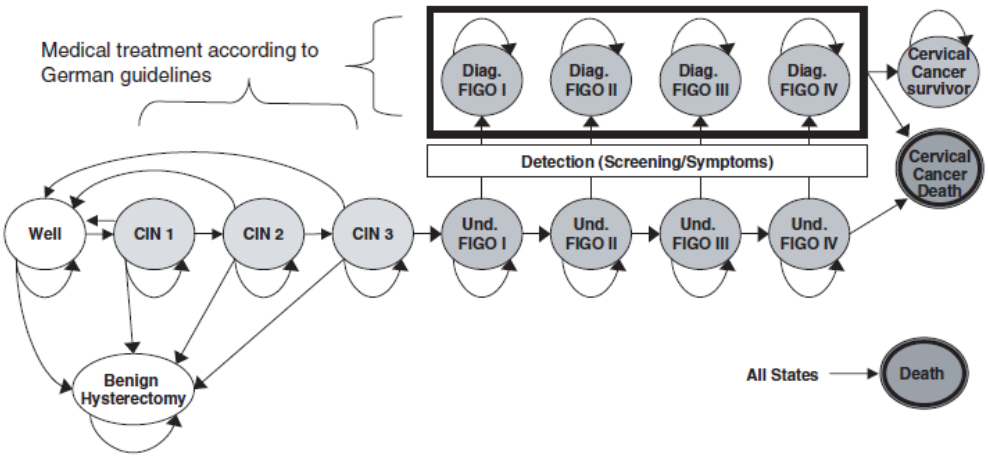

Figure 2: Schematic diagram of a Markov model used as the German cervical cancer screening model. (adapted from (37)) A hypothetical cohort of women may acquire a different health state within a lifetime based on cervical cytology/histology: no cervical lesion (well), benign hysterectomy (Benign hysterectomy), mild cervical intraepithelial neoplasia (CIN 1), moderate cervical intraepithelial neoplasia (CIN 2), cervical intraepithelial neoplasia/carcinoma in situ (CIN 3), undiagnosed invasive cervical cancer FIGO states I–IV (Und. FIGO I–IV), diagnosed invasive cervical cancer FIGO states I–IV (Diag. FIGO I–IV), cervical cancer survivors 5 years after cervical cancer diagnosis and treatment (Cervical Cancer Survivor), death from cervical cancer (Cervical Cancer Death) and death from other causes (Death). Women may remain in the same health state, progress or regress to another health state, may die from cervical cancer as a function of FIGO-specific survival rates or may die from other causes as a function of age and gender.

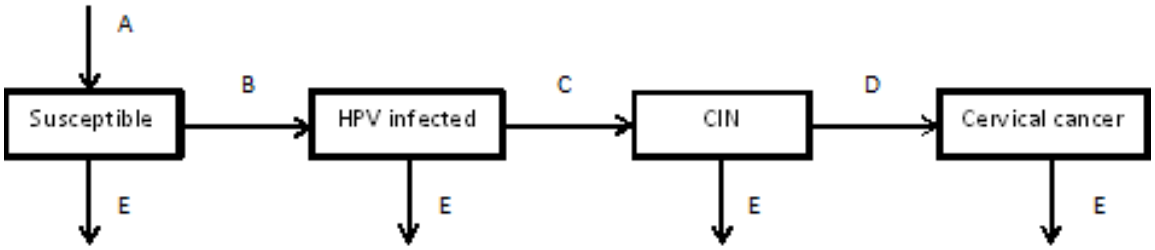

Figure 3: Schematic diagram of a dynamic model (adapted from (38)). The entry of new susceptible individuals into the model is represented by the parameter A (people being born). Individuals exit the model at the mortality rate E, which can be compartment specific. Parameter B represents the rate at which susceptible individuals become infected over a small period of time (parameter B changes with time, hence reducing the prevalence of HPV infection over time means that susceptible individuals are less likely to become infected because there are fewer persons in the population to infect them with HPV – this indirect effect of vaccination is called 'herd immunity effect'). To account for the changing prevalence of HPV infection in the population, the parameter B is measured as a function of time, age, the number of sexually active persons in the population who are infected and not infected, the way they form sexual partnerships, and the transmission probability of HPV infection per partnership. The parameter B thus accounts for the transmission dynamics of HPV infection over time. Parameters C and D are equivalent to parameters (probability of progressing to another state of the disease) in a cohort model.

2. AIM

The objective of the thesis is to describe issues around the transferability of health economic evaluations, and to test one of the recently published methods for assessing transferability, the EURONHEED checklist. For the purpose of testing the method, whose concept was published by Boulenger (2005) and supplemented in detail by Nixon (2009), an economic evaluation of HPV vaccination was chosen as a study case.

3. METHODS

First, a systematic search of published economic evaluations of HPV vaccination was performed. Based on the pre-defined inclusion criteria appropriate studies were selected and scored using the EURONHEED checklist.

3.1. Systematic literature search

Published studies were searched in two for economic evaluations crucial databases: MEDLINE and NHS EED databases.

MEDLINE (Medical Literature Analysis and Retrieval System Online) is a bibliographic database of life sciences and biomedical information. It includes bibliographic information for articles from academic journals covering medicine, nursing, pharmacy, dentistry, veterinary medicine, and health care. MEDLINE database is freely available on the internet and searchable via PubMed (http://www.ncbi.nlm.nih.gov/pubmed/).

The NHS Economic Evaluation Database (NHS EED) is a collection of critical assessments of published economic evaluations of health care interventions. The purpose of the database is to assist researchers and decision makers in systematically identifying, interpreting, appraising the quality of economic evaluations, which are spread over many databases and paper based resources. NHS EED is available free of charge on the internet (http://www.york.ac.uk/inst/crd/).

The search was performed in July 2010. In MEDLINE the following keywords or medical subject headings were used: "*costs and cost analysis"* [Mesh], "*Papillomavirus vaccines*" [Mesh], and "*HPV and vaccin**". The search was limited to non-review studies not older than ten years, published in the English language, and concerning human only. The search profile as used in MEDLINE is presented in Table IV.

In NHS EED database the following searching criteria was used: 'HPV AND economic evaluation AND vaccine'.

Table IV: Search profile in MEDLINE

	Search criteria
#1	"Costs and Cost Analysis"[Mesh]
#2	"Papillomavirus Vaccines"[Mesh]
#3	#1 and #2
#4	#1 and (#2 or (HPV and vaccin*))
#5	#4 Limits: Humans, English, published in the last 10 years, Field: Publication Type
#6	#5 NOT review [Publication Type] Limits: Humans, English, published in the last 10 years Field: Publication Type

3.2. Inclusion criteria

For a study to be included in the transferability assessment the following inclusion criteria were determined:

1. Study performed in a country with an advanced economy based on the International Monetary Foundation (IMF) classification.
2. Study is a full economic evaluation
3. Study assessed HPV vaccine alongside cervical cancer screening versus screening alone
4. Either bivalent or quadrivalent HPV vaccine was evaluated
5. HPV vaccination of girls only in pre-sexual period

Firstly, we included studies performed in countries on the IMF advanced economies list as these countries have a comparably developed health care system. Secondly, studies had to evaluate both costs and health benefits in order to be treated as an economic evaluation. Thirdly, studies which assessed the introduction of HPV vaccine alongside cervical cancer screening were selected as this is the way that HPV vaccination is applied in Slovenia and most of the other countries. Fourthly, since there are both a bivalent and a quadrivalent vaccine registered in Europe studies evaluating either of them were included. Lastly, studies evaluating the vaccination of girls in pre-sexual period were selected as this target group is represents the population that is currently vaccinated.

3.3. EURONHEED checklist

As mentioned in the introduction, Nixon et al. published a supplemented version of EURONHEED transferability evaluation checklist (20), which was first time introduced by Boulenger et al. in 2005.(8) This checklist represents a part of collaborative research between centres participating in the European Network of Health Economic Evaluation Databases (EURONHEED) project. It was developed as a method for evaluating the transferability and generalizability of studies to be considered for inclusion in databases that comprise the EURONHEED network.

Since the first publication of the checklist the correspondence with researchers has indicated the need for further clarifications in completing the checklist. Therefore Nixon et al. provided more details about the each of 42 items so the checklist can be more consistently completed (Table V). They wanted to ensure more clarity in the interpretation of what is being assessed in each of the checklist items. The scoring process is, however, the same as described in the introduction (see page 22).

Table V: EURONHEED transferability information checklist with Nixon's instructions (items marked with [a] belong to the transferability subchecklist)

Study question	**Q1**	**Is the study question clearly stated?**
	Did the author(s) clearly state the hypothesis/aim/ objective of the study? Did they fully describe what they were trying to demonstrate?	
	Q2	**Are the alternative technologies justified by the author(s)?**
	Did the authors clearly explain the reason(s) for their choice of comparator(s)? If they simply stated their choice of comparator(s) the answer should be 'no'. If they provided a description of potential comparators and a rationale for their selection, or if they chose current practice, the answer should be 'yes'.	
Health technology	**HT1[a]**	**Is the intervention described in sufficient detail?**
	The principal question being addressed is whether or not enough information is provided for a reader to be able to fully understand the health technology studied, and whether or not it is applicable to other contexts. To provide a 'yes' answer the article should provide a full description of the technology studied, which in the case, for example, of pharmaceuticals includes dosages, relevant intervals, method of administration (for example, orally, subcutaneously), information on how many times per hour/day/week, for how long, and where and by whom it was administered. If only a general description is given the answer should be 'partially', or 'no' if only the name of the health technology is given.	
	HT2[a]	**Is (are) the comparator(s) described in sufficient details?**
	See explanation above for HT1.	
Setting	**SE1**	**Did the authors correctly specify the setting in which the study took place (e.g. primary care, community)?**
	Give responses according to the five cases possible: correctly specified: answer 'yes'; misspecified but can be inferred: answer 'partially'; misspecified but cannot be inferred: answer 'no'; not specified but can be inferred: answer 'partially'; not specified but cannot be inferred: answer 'no'.	
	SE2[a]	**Is (are) the country state which the economic study took place clearly specified?**
	Answer either 'yes' or 'no' as appropriate.	
Perspective	**P1[a]**	**Did the authors correctly state which perspective they adopted for the economic analysis?**
	This question reflects the plausible scenarios with regard to the economic perspective (stated or not) and whether or not the cost analysis matched the stated/inferred perspective; for example, if the authors stated a societal perspective but did not include indirect costs (productivity changes), the response should be 'partially'. Give responses according to the five cases possible: if correctly specified with all relevant costs included: answer 'yes'; if misspecified but can be inferred from the reporting of costs: answer 'partially'; if misspecified but cannot be inferred from the reporting of costs: answer 'no'; if not specified but can be inferred from the reporting of costs: answer 'partially', if not specified and cannot be inferred from the reporting of costs: answer 'no'.	

Study population	**SP1[a]**	**Is the target population of the health technology clearly stated by the authors or when it is not done can it be inferred by reading the article?**
	The target population is the entire group a researcher is interested in, the group about which the researcher wishes to draw conclusions. This may differ from the actual study population and therefore affect the generalizability of the results.	
	SP2	**Are the population characteristics described? (e.g. age, sex, health status, socio-economic status, inclusion/exclusion criteria)**
	This question refers to the population from which the sample was drawn. To answer 'yes' the author(s) need(s) to have described the demographic profile and disease status where relevant of the studied population. If only a general description, i.e. a description not or only partially allowing the reader to assess if it is applicable to the population of interest is given, the answer should be 'partially', or 'no' if the study population is not described.	
	SP3[a]	**Does the article provide sufficient detail about the study sample(s)?**
	To answer 'yes' the authors need to have reported all of the following: whether the sample size was determined prior to the selection process and if so, how; the sample selection process; sample inclusion/exclusion criteria (which may be additional to study population inclusion/exclusion criteria); the percentage of patients who refused to participate or who were excluded; the number of subjects in each group; the characteristics of the subjects. If only some/none of the above items are reported, answer 'partially' or 'no'. If the clinical data are derived from a literature review, there may be many samples from many studies and this detail on the samples is unlikely to be reported, in which case the answer should be 'not applicable'.	
	SP4	**Does the paper provide sufficient information to assess the representativeness of the study sample with the respect to the target population?**
	If the subjects in the study sample have characteristics different from those of the target/study population, these differences must be considered to be potential confounders and may bias the outcome of the study. To answer 'yes' the author(s) need(s) to have described the sample in sufficient detail and when and if it differs from the study population inclusion/exclusion criteria. They may also compare the characteristics, where relevant, of those who refused to be included in the study with those who participated, and determine if people who dropped out of the study were different in comparison with those who remained (if the study participants were found to be different from those who were not in the study, the results could be biased because the subjects who were potentially the 'worst cases' were not included in the study). If the authors did only some of the above the answer should be 'partially'. If the clinical data are derived from a literature review and there are many samples from many studies, the above guidelines still apply.	
Modelling	**M1**	**If a model is used is it described in detail?**
	When a model is used, to answer 'yes' the authors should have clearly stated the purpose of the model, described the type of model used (e.g. decision tree, Markov model), the key assumptions of the model, provided details of the software used, and the time horizon the model is examining. 'Partially' should be answered only if some of these requirements 'yes' are provided, and 'no' if no description is made of the model (stating that a decision tree, for example, is used, is not enough). If no model is used, or other modelling such as regression analysis is performed, write 'not applicable'.	
	M2	**Are the origins of the parameters used in the model given?**
	To answer 'yes' the authors need to have provided detailed descriptions of the sources (e.g. literature, expert opinion, single trial) used to derive the model's parameters (point estimates and ranges if appropriate). 'Partially' should be the answer if only some of the sources are given. Choose 'no' if no sources are given.	

Effectiveness	**E1**	**If a single study is used is the study design described (sample selection, study design, allocation, follow-up)?**
	To answer yes the authors need to have described the method of sample selection, the study design, if the study was single or multicentred, the allocation method, the duration of follow-up and the loss to follow-up (if relevant). 'Partially' should be the answer if any of the requirements for 'yes' are not met. 'No' should be the answer if either the type of study, whether or not it was single or multicentred or the duration of follow-up was not mentioned.	
	E2	**If a single study is used are the methods of data analysis described (ITT/per protocol or observational data)?**
	Did the authors: state, in the case of an experimental study, if the study was based on intention-to-treat or treatment completers, and in the case of an observational study, if all patients included in the study were accounted for in the analysis; report the primary health outcomes used in the analysis and any particular instrument used to measure them; discuss the comparability of the groups. If only some/none of these details were given answer 'partially'/'no.'	
	E3	**If based on a review/synthesis of previous published studies, are review methods described (search strategy, inclusion criteria, sources, judgment criteria, combination, investigation of differences)?**
	There are many important steps in a review. However, it is recommended that to answer 'yes' the authors should undertake the following: state the inclusion and/or exclusion criteria, list the sources searched by (e.g. MEDLINE, unpublished data), specify the methods used to combine the results of the individual primary studies (if applicable).If only some/none of these details were given answer 'partially' or 'no'.	
	E4	**If based on opinion, are the methods used to derive estimates described?**
	To answer 'yes' the authors should have reported the methods used to derive estimates of effectiveness (model parameters), e.g. consensus, experts' opinion and authors' assumptions.	
	E5[a]	**Have the principal estimates of effectiveness measures been reported?**
	To answer 'yes' the authors need to have fully provided (in terms of principal effectiveness measures) the results of the clinical trial, the study, the literature review (or parameters used in the model) or the results based on opinion.	
	E6	**Are the side effects or adverse effects addressed in the analysis?**
	To answer 'yes' the author(s) should provide quantitative results relating to side-effects or adverse events. However, if the health technology being studied is not associated with side-effects (although most interventions have some degree of side-effect) the answer should be 'not applicable'.	
	E7[a]	**Does the article provide the results of a statistical analysis of the effectiveness results?**
	Did the authors present the 95% confidence intervals and/or the p values. The responses to this question will normally be 'yes,' 'no' or 'partially' depending on the level of reporting. This item would not normally apply to modelling studies.	

Benefit measure	**B1**	**Do the authors specify any summary benefit measure(s) used in the economic analysis?**
	Is a measure of benefit used for the economic analysis? For example, lives saved, numbers of life years gained, or quality-adjusted life years (QALY). If a cost-consequences analysis is performed (e.g. several clinical outcomes are reported as in the case of a surgical intervention) write 'not applicable'.	
	B2	**Do the authors report the basic method of valuation of health states or interventions?**
	This question only applies to cost-utility and cost-benefit analyses. In the case of a cost-effectiveness study (including cost-consequences or cost-minimization studies), write 'not applicable'. To answer 'yes' the author(s) should report the method of valuation of health states (using generic or healthspecific valuation tools), willingness-to-pay, human capital, etc.	
	B3	**Do the authors specify the source(s) of health states (e.g. Specific patient population or the general public)?**
	Answer this question only if B2 is relevant. Answer 'yes' if the authors indicate whose values were used to assess health states: e.g. authors' assumption, clinician and patients.	
	B4	**Do the authors specify the valuation tool used?**
	Answer this question only if you answered 'yes' to B2. To answer 'yes' the authors need to specify standard gamble, time trade-off, conjoint analysis, etc. If a standard generic measure is used (e.g. EQ-5D, Health Utility Index (HUI), SF-6D) write 'not applicable'.	
	B5[a]	**Is the level of reporting of benefit data adequate (incremental analysis, statistical analyses)?**
	To answer 'yes' the author(s) need to have reported the appropriate results for each study subgroup, presented the incremental results (when applicable), and the results of any statistical tests. The response should be 'not applicable' if the answer to question B1 was 'no' or 'not applicable'.	
Costs	**C1[a]**	**Are the cost components/items used in the economic analysis presented?**
	To answer 'yes' the author(s) need(s) to have stated which costs (drug, personnel, etc.) they measured and which costs were included in the final cost figure.	
	C2	**Are the methods used to measure costs components/items provided?**
	This question requires an assessment of the way each cost item was calculated. For example, was microcosting conducted or were diagnostic-related group costing used? Irrespective of the approach used, it should be fully described. Answer 'yes' if the methods are fully provided, 'partially' if methods are provided for some costs only or if the methods are not sufficiently described,'no' when no information is provided about the method of calculation of the cost items.	
	C3	**Are the sources of resource consumption data provided?**
	Answer 'yes' if the authors provide full details of sources, which may be: prospective or retrospective study (actual data), a model, a literature review, Health Department data, etc. Answer 'partially' if sources are provided for some costs only.	
	C4	**Are the sources of unit price data provided?**
	Answer 'yes' if the author(s) detail where their unit prices come from, e.g. hospital source, published literature, or official prices, Answer 'partially' if only some sources are provided.	
	C5[a]	**Are unit prices for resources given?**
	Answer 'yes' if the authors give the unit price for each resource consumption item; 'partially' if only some unit prices are provided.	
	C6[a]	**Are costs and quantities reported separately?**
	Answer 'yes' if the author(s) provide, for each resource included, individual costs and their associated quantities used, Answer 'partially' if some costs and quantities were reported separately.	
	C7[a]	**Is the price year given?**

Costs	*Answer 'yes' if the cost data are presented (possibly related if resource consumption relates to different years) for a given price year.*	
	C8	**Is the time horizon given for each element of the cost analysis?**
	Here we want to know the period of time covered by the cost measured (e.g.: (1) the cost of rehabilitation covers a 6-month period, which corresponds to the length of follow-up, (2) the cost of drugs were measured for the lifetime of a patient). Answer 'yes' if the time horizon for each cost element is known. This is important as it will enable judgment regarding the need for discounting.	
	C9[a]	**Is the currency unit reported?**
	Answer 'yes' if the currency is provided.	
	C10	**Is a currency conversion rate given?**
	This question only applies to studies in which the results were converted from one currency to another. Answer 'yes' if the results were converted and the conversion rate given. This question is mainly applicable to multicountry studies when the cost figures for each country are converted into a single currency unit (in which case all conversion rates should be reported). Also applicable to single-country studies when, for example, a French study is published in an American journal and the cost data (in Euros or in Francs) are converted into American dollars. The method should also be given (e.g. exchange rate or Purchasing Power Parities, PPP). Answer 'partially' if the rate is provided for some countries in the analysis but not for all; answer 'not applicable' if no conversion of the results or data was performed.	
	C11	**Does the article provide the results of a statistical analysis of cost results?**
	Answer 'yes' if the quantities/cost data were treated stochastically and appropriate measures of precision (e.g. p values) given. Answer 'partially' if descriptive statistics were provided (such as mean and standard deviation), and 'no' if treated deterministically (i.e. only point estimates given). If a model is used, answer 'yes' if the methods to deal with cost uncertainty, such as sensitivity analyses or probabilistic sensitivity analyses, are provided.	
Discounting	**D1**	**Was the summary benefit measure(s) discounted?**
	Answer 'yes' if the time horizon warranted discounting and it was undertaken. Answer 'no' when the time horizon warranted discounting but it was not undertaken. Answer 'not applicable' when the time horizon is below 1 year.	
	D2	**Were the costs data discounted?**
	The same as for D1.	
	D3	**Do the authors specify the rate(s) used in discounting costs and benefits?**
	Answer 'yes' if these data were given and relevant; 'not applicable' when the time horizon did not warrant discounting for costs and benefits.	
	D4	**Were discounted and not discounted results reported?**
	Answer 'not applicable' if discounting for costs and benefits was not relevant.	
Discussion	**S1[a]**	**Are quantitative and/or descriptive analysis conducted to explore variability from place to place?**
	Answer 'yes' if the authors assessed quantitative variability in the data through (for example) sensitivity analysis (e.g. modelling using country-specific data in multinational studies, or applying country-specific cost data to determine results). Answer 'partially' or 'no' if descriptive or no comments (variations in practice) were given.	
	O1[a]	**Did the authors discuss caveats regarding the generalizability of their results?**
	Answer 'yes' if, in the discussion of the paper, the authors undertook an appraisal of how the particular features and methods of their study may limit the relevance of their findings to other locations or countries.	

While scoring the included studies we realised that the general guidance for assessing items sometimes allows a too subjective interpretation of the question. Since it was the purpose to answer questions following the same criteria each time we decided to additionally specify some of the items.

For answering the question HT1 with 'yes' coverage rate of girls participating vaccination programme and the age of vaccinated girls had to be reported (the vaccination schedule was the same for every setting, i.e. applying 3 vaccine doses). To answer 'yes' on question HT2 authors had to provide the following data about PAP smear screening: which age group participates screening, how frequently is screening performed, screening pattern (clinical practice after a positive result of the screening), sensitivity and specificity of the test, and screening coverage rate. If only some of information was provided, the answer was 'partially'. To answer with 'yes' on item SP_1 (study population) we expected the age of the girls vaccinated to be reported, since we did not have a clinical trial population sample in our case but model based studies where scenario of the population was simulated.

4. RESULTS

4.1. Results of search and literature review

The search criteria resulted in 129 and 30 hits in MEDLINE and NHSEED databases, respectively. Based on the review of authors and titles 18 studies were excluded due to duplication. The remaining 141 abstracts were reviewed to check if the study met the inclusion criteria. As soon as one of the inclusion criteria was not fulfilled the study was excluded. Sometimes more than one inclusion criterion was not met. Most of the studies (107) were excluded after reading the abstract. Additional 4 studies were excluded after reading the full text.

A total number of articles excluded according to each inclusion criterion are presented in Table VI. Most of excluded articles were not fulfilling the criterion of being a full economic evaluation. These articles were either reviews, experts' opinions or commentaries. In the end, 30 of the 159 studies found were scored. A summary of the included studies is presented in Table VII. According to inclusion criteria all identified studies included into scoring were model-based studies.

Table VI: Excluded studies based on reviewing the abstract or full text.

Reason for not including the study	No. of studies
Study was not performed for the developed country (IMF classification)	**18**
Study was not economic evaluation	**71**
Evaluated technologies were not HPV vaccine alongside screening versus screening alone	**4**
Study was not assessing cost and effects of vaccination only of girls in pre-sexual period	**5**

Table VII: Summary of studies included for scoring.

No.	Author (year)	Country	Ref.	Perspective	Type of model
1.	Annemans (2009)	Belgium	(39)	Belgian health care payer	Markov model
2.	Bergeron (2008)	France	(40)	Two perspectives: direct health care cost perspective (payer and patients) and health care payer perspective	Markov model
3.	Boot (2007)	Netherlands	(41)	Not specified	Markov model
4.	Brisson (2007)	Canada	(42)	Ministry of health (direct medical costs)	Markov model
5.	Chesson (2008)	USA	(43)	Social perspective and direct medical costs	Markov model
6.	Coupe (2009)	Netherlands	(44)	Not specified	Markov model
7.	Dasbach (2008)	Taiwan	(45)	Health care system perspective	Dynamic model
8.	de Kook (2009)	Netherlands	(46)	Societal perspective (indirect costs of vacc. and screening are included)	Dynamic model
9.	Elbasha (2007)	USA	(47)	US health care system	Dynamic model
10.	Ginsberg (2009)	Multiregional	(48)	Not specified	Markov model
11.	Ginsberg (2007)	Israel	(49)	Health care system perspective	Markov model
12.	Goldahaber-Fiebert (2008)	USA	(50)	Societal perspective	Microsimulation model
13.	Goldie (2004)	USA	(51)	Societal perspective	Markov model
14.	Jit (2008)	UK	(52)	Health care provider perspective	Dynamic model
15.	Kim (2008)	USA	(53)	Societal perspective, but only direct medical and non-medical costs included	Dynamic model
16.	Kulasingam (2007)	Australia	(54)	Government perspective	Markov model
17.	Kulasingam (2003)	USA	(55)	Not specified	Markov model
18.	Mennini (2008)	Italy	(56)	Health care provider perspective	Markov model
19.	Oddsson (2009)	Iceland	(57)	Not specified	Simplified economic calculation base on CIN and ICC reduction
20.	Rogoza (2009)	Netherlands	(58)	Not clearly specified	Markov model
21.	Sanders (2003)	USA	(59)	Not specified	Markov model
22.	Szucs (2008)	Switzerland	(60)	Direct health care cost perspective	Markov model
23.	Thiry (2009)	Belgium	(61)	Belgian health care payer	Markov model
24.	Usher (2008)	Ireland	(62)	Health care payer	Dynamic model
25.	Zechmeister (2009)	Austria	(63)	Health care payer and social perspective	Dynamic model
26.	Dasbach (2008)	Norway	(64)	Norwegian health care system	Dynamic model
27.	Kulasingam (2008)	UK	(65)	Health system perspective	Markov model
28.	Rogoza (2008)	Multiregional	(66)	Health care payer and societal perspective	Markov model
29.	Suarez (2008)	Multiregional	(67)	Health care payer	Markov model
30.	Taira (2004)	USA	(68)	Not specified	Dynamic model

4.2. Results of scoring

Scores for each study and item are shown in Table VIII (table spans across two pages). At the bottom of each column the final score represents the percentage of the maximum score.

The minimal score calculated in the checklist was 23.3 (subchecklist 41.7) and the maximal score 79.0 (subchecklist 100.0). The average score was 61.7 (subchecklist 75.3), and the median 63.5 (subchecklist 79.2). The standard deviation was 12.6 (subchecklist 29.4).

Figure 4 shows the distribution of total scores across different ranges for the checklist and the subchecklist separately. Most of the studies (40%) achieved 60-70% of the maximum score. Scores were higher if only items forming the subchecklist were selected, in which most (40%) studies having 80-90% of the maximum score.

The correlation between results of the checklist and subchecklist is presented in Figure 5. The correlation was strong, with a Pearson coefficient R^2 equal to 0.74.

Table VIII: Results of scoring of 42 checklist items for each study. At the bottom scores of checklist and subchecklist are summarized using the formula presented in Methods.

Checklist items	*Sequential number of the study included in scoring*																													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Q1	1.0	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Q2	1.0	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0
HT1[a]	**1.0**	**1.0**	**0.5**	**0.5**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**0.5**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
HT2[a]	**1.0**	**1.0**	**0.5**	**0.0**	**0.0**	**1.0**	**0.0**	**1.0**	**0.5**	**1.0**	**1.0**	**1.0**	**1.0**	**0.0**	**1.0**	**1.0**	**1.0**	**0.5**	**0.0**	**0.5**	**1.0**	**1.0**	**0.5**	**1.0**	**0.5**	**0.5**	**1.0**	**1.0**	**1.0**	**0.0**
SE1	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
SE2[a]	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
P1[a]	**1.0**	**1.0**	**0.0**	**1.0**	**1.0**	**0.0**	**1.0**	**1.0**	**1.0**	**0.0**	**1.0**	**1.0**	**1.0**	**1.0**	**0.5**	**1.0**	**0.5**	**1.0**	**1.0**	**1.0**	**0.5**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**0.0**
SP1[a]	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
SP2	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
SP3[a]	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**
SP4	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
M1	1.0	1.0	0.0	0.0	1.0	1.0	0.5	1.0	1.0	0.0	1.0	0.5	1.0	0.5	1.0	1.0	0.5	0.5	1.0	0.5	1.0	1.0	0.5	1.0	1.0	0.5	1.0	0.5	0.5	1.0
M2	1.0	1.0	0.0	0.5	1.0	1.0	0.0	0.5	1.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	0.5	1.0	1.0	0.0	1.0	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	0.0
E1	0.0	0.0	0.0	0.0	0.0	N	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
E2	0.0	0.0	0.0	0.0	0.0	N	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
E3	N	N	N	0.0	N	0.0	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
E4	N	N	N	0.0	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
E5[a]	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**1.0**	**0.0**	**0.0**	**0.0**	**0.0**	**1.0**	**1.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**1.0**	**1.0**	**0.0**	**0.0**	**1.0**	**0.0**	**0.0**	**0.0**
E6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	N	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
E7[a]	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**

Checklist items	*Sequential number of the study included in scoring*																														
B1	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
B2	1.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	N	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.5	N	N	1.0	0.0	0.0	0.0	0.0
B3	1.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	N	0.0	1.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	N	N	1.0	0.0	0.0	0.0	0.0
B4	0.0	0.0	N	N	N	N	N	N	N	N	N	N	N	N	0.0	N	0.0	N	N	N	0.0	N	0.0	N	N	0.0	N	0.0	0.0	N
B5[a]	**1.0**	**1.0**	**0.0**	**1.0**	**1.0**	**1.0**	**1.0**	**0.0**	**0.0**	**0.5**	**1.0**	**1.0**	**1.0**	**0.0**	**0.0**	**0.0**	**1.0**	**1.0**	**0.0**	**0.0**	**1.0**	**1.0**	**0.5**	**0.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
C1[a]	**1.0**	**1.0**	**1.0**	**1.0**	**0.0**	**1.0**	**1.0**	**1.0**	**0.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**0.0**
C2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
C3	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.5	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0
C4	1.0	1.0	0.0	0.5	0.0	0.5	0.0	0.5	0.0	0.0	1.0	0.5	1.0	1.0	0.5	1.0	0.5	1.0	0.0	0.0	1.0	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	0.0
C5[a]	**1.0**	**1.0**	**0.0**	**1.0**	**0.0**	**1.0**	**0.0**	**1.0**	**0.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**0.5**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**0.0**
C6[a]	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**
C7[a]	**1.0**	**1.0**	**0.0**	**1.0**	**0.0**	**1.0**	**1.0**	**0.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**0.0**	**1.0**	**1.0**	**1.0**	**0.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**0.0**
C8	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
C9[a]	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
C10	N	N	N	N	N	N	N	N	N	0.0	1.0	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	1.0	N	0.0	0.0	N
C11	1.0	1.0	0.0	0.5	1.0	0.0	1.0	0.5	0.5	0.0	0.5	1.0	1.0	1.0	0.0	0.5	1.0	1.0	0.0	0.5	1.0	1.0	0.5	0.5	1.0	0.5	1.0	1.0	0.0	0.0
D1	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0
D2	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0
D3	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0
D4	1.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	0.0	0.5	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	1.0	0.5	1.0	0.0	1.0	0.5	1.0	0.0
S1[a]	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**1.0**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**0.0**	**0.0**	**N**
O1[a]	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**1.0**	**0.0**	**1.0**	**0.0**	**1.0**	**1.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.5**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**1.0**	**1.0**	**1.0**
checklist score (%)	**77.4**	**74.2**	**23.3**	**56.3**	**53.3**	**67.2**	**55.0**	**61.7**	**51.7**	**56.7**	**79.0**	**63.3**	**66.7**	**61.7**	**53.2**	**68.3**	**64.5**	**66.7**	**46.7**	**55.0**	**75.0**	**70.0**	**62.9**	**71.4**	**73.2**	**64.1**	**76.7**	**63.6**	**62.1**	**30.0**
subchecklist score (%)	**83.3**	**83.3**	**41.7**	**70.8**	**50.0**	**83.3**	**66.7**	**75.0**	**54.2**	**76.9**	**100.0**	**83.3**	**83.3**	**66.7**	**66.7**	**83.3**	**87.5**	**70.8**	**66.7**	**70.8**	**83.3**	**75.0**	**83.3**	**83.3**	**79.2**	**79.2**	**91.7**	**84.6**	**84.6**	**50.0**

[a] this item forms the transferability subchecklist
N –'not applicable'.

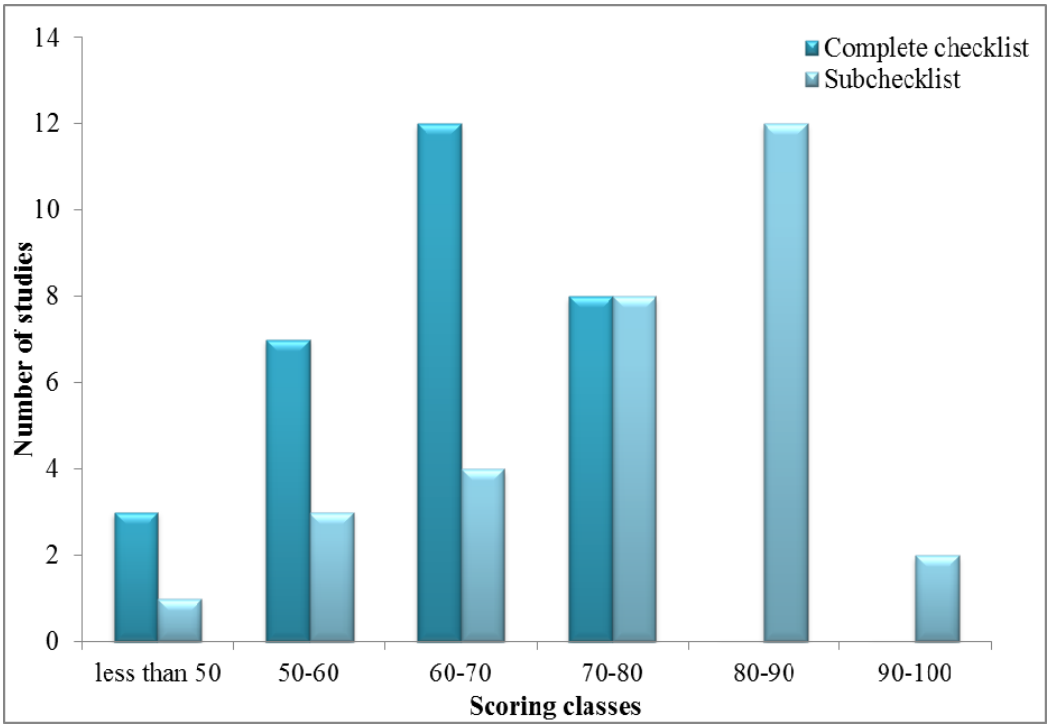

Figure 4: Distribution of studies across scoring ranges.

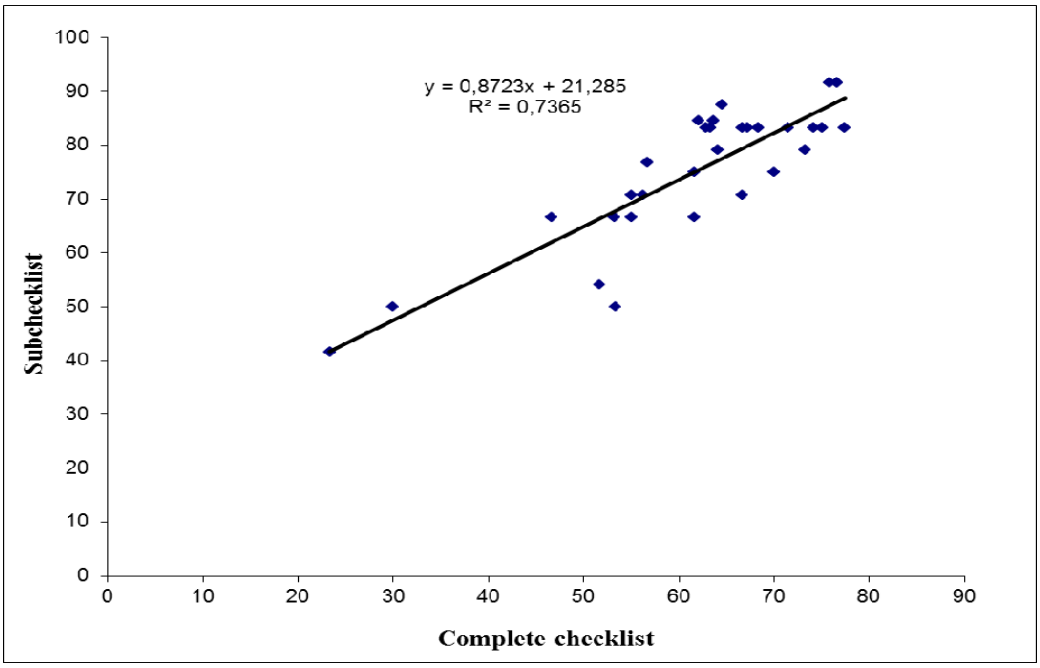

Figure 5: Correlation between subchecklist scoring and scoring with complete checklist.

Figure 6 illustrates how particular items were scored across all studies, shown as a percentage of the maximum score possible. Only items that were applicable in at least one study are shown. Items SE_1, $SP_{2\text{-}4}$, E_7, C_6, and C_8 were not applicable as all studies were model-based economic evaluations. The extent to which different items were applicable for our 30 studies is shown in Figure 7.

Figure 8 is similar to Figure 6 but it shows how a particular group of items scored across studies. Percentages of the total score for items forming the same group were averaged out.

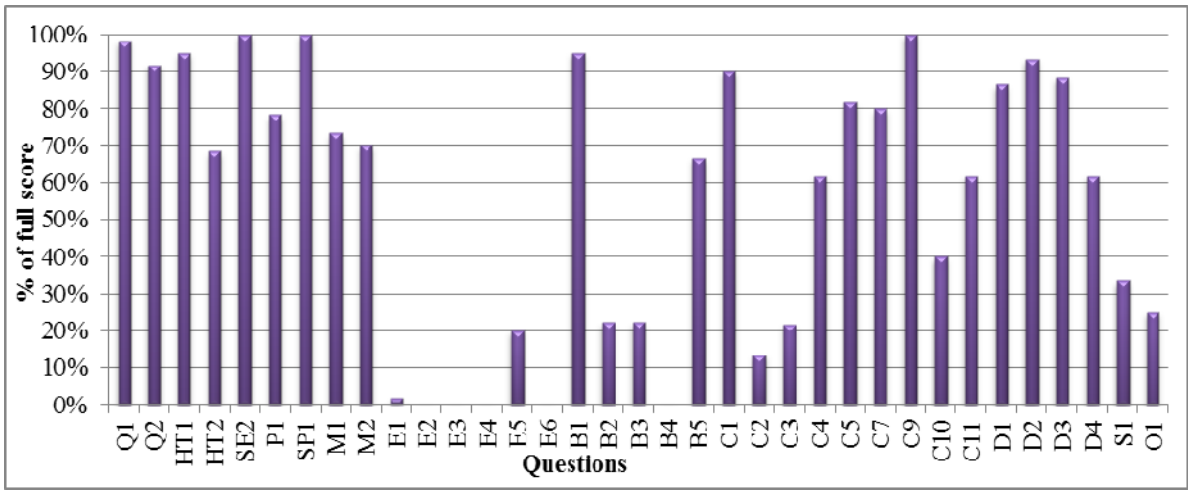

Figure 6: Percentage of the total score achieved for each item across all studies.

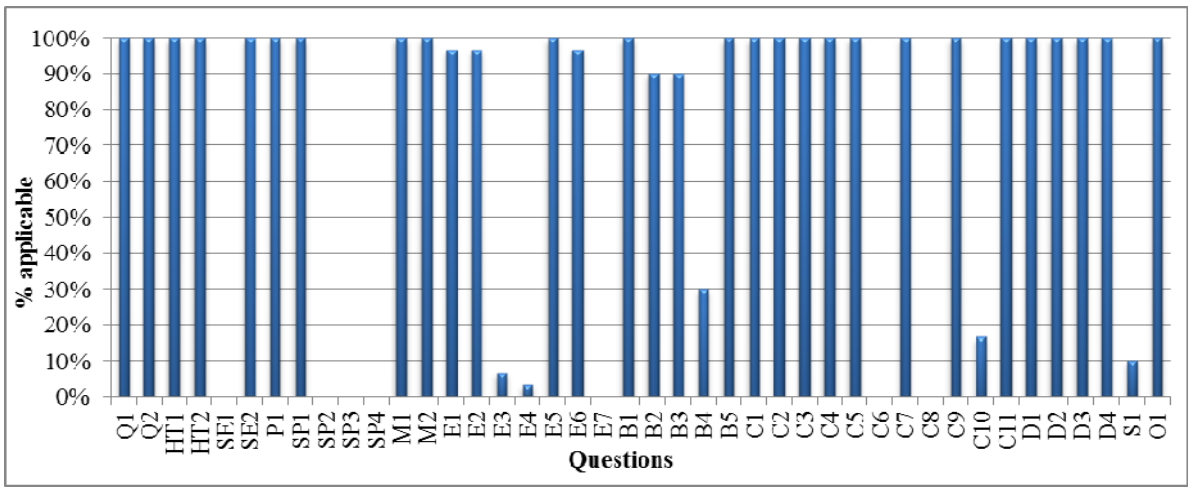

Figure 7: Applicability of items across all studies.

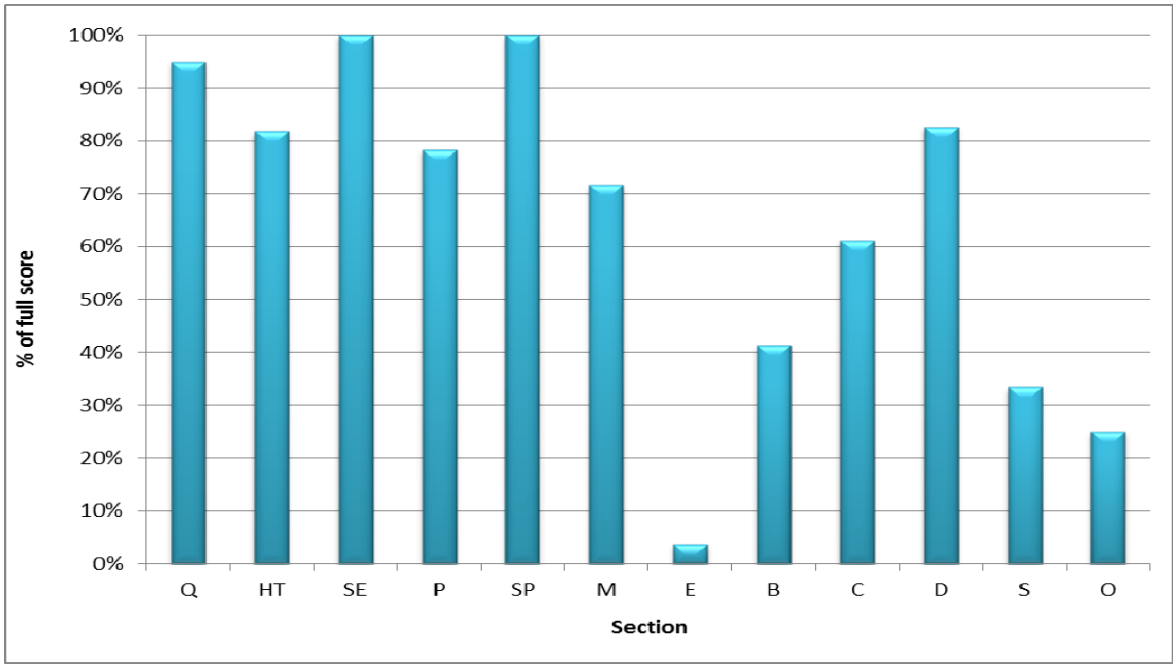

Figure 8: Percentage of scores achieved inside of each group of items. Q – Study question, HT - Health technology, SE - Setting, P - Perspective, SP – Study population, M – Modelling , E - Effectiveness, B – Benefit measure, C - Costs, D - Discounting, S – sensitivity analysis exploring variability from place to place, O – authors' opinion about generalizability of results.

5. DISCUSSION

In our research work we have tested the performance of a recently published checklist for assessing the transferability of economic studies which is called the EURONHEED checklist. As a case study economic evaluations of HPV vaccination of girls only in pre-sexual period alongside cervical cancer screening versus screening alone were chosen. In the following paragraphs the method used and the results obtained are first discussed. This is followed by comparison to other methods and discussion about how to improve the transferability of economic evaluations.

5.1. Comment on method used for scoring

All studies were model-based studies. The methodology of model-based economic studies is different from economic evaluations performed alongside clinical trials, which means that the usefulness of the checklist may be different for these two different study types. We have experienced that additional instructions for scoring 42 items of the checklist (published by Nixon et. al.) were helpful only to some extend and still allow quite subjective interpretation of the questions. As a result, scoring results might thus be dependent on the assessor. To avoid this deficiency of the method we additionally specified some of the items (HT_1, HT_2 and SP_1) as described in the Methods section.

The section of the checklist *Health technologies* (questions HT_{1-2}) was one of the most difficult to assess due to the nature of our interventions compared – HPV vaccination with screening versus screening alone. Instructions how to assess items HT_{1-2} are more suitable for intervention in health care such as oral pharmaceuticals administration. Therefore, the questions about dosages, methods of administration, relevant intervals etc. were not well applicable in our case. We decided to define our own criteria, which we asssumed to be the most important for decision makers to be reported. Since the vaccination schedule is always the same, we were interested to know what kind of coverage rate for the vaccination authors predicted in their setting and the age of girls being vaccinated. We expected that the description of the clinical practice of screening will provide the following data: age group of those who participate in the screening, screening coverage rate, the frequency of screening, screening pattern (clinical practice after a positive result of the screening) and sensitivity and specificity of the test. If only some of information was

provided, the answer was 'partially' and only a score of 0.5 was given. We decided to specify these criteria as we expected that these factors influence the final result of the cost-effectiveness of the study and consequently the transferability of the study.

Item SE_1 asks about reporting the setting of the study (primary care, secondary care, etc.). We did not count this item as applicable for our example of vaccination, although this scoring option is not specifically mentioned in the instructions for scoring.

The section about *Study population* (SP_{1-4}) refers primarily to clinical trial based economic evaluations, where a sample from the population is taken and the data about the sample (size, selection process, exclusion/inclusion of patients) need to be reported. In our case we expected that the age of the girls vaccinated was reported for item SP_1 as the most relevant characteristic of targeted population and it was shown to be stated in every published article. The rest of the items, referring to specific characteristics of study population and study sample were assessed as 'not applicable'.

The section about modelling surprisingly contains only two questions. This gives small relative weight to this section compared to other sections. It was noticed while scoring that the level of reporting details about modelling varied a lot among studies, although owing to loosely specified criteria the scores given were similar. It happened that two compared studies both gained same value of scores for M_1 item, but one provided much more details than another. According to the definition of transferability by the authors of the EURONHEED checklist (8), which stated that "data, methods and results are transferable if a potential user can assess their applicability to their setting and they are applicable to that setting", potential user needs to be well-informed about the performed method described in a study and needs to be provided with the detailed information about the method, input data, and how the results were gained.

The effectiveness section consists of seven items where the first four (E_{1-4}) items relate to the sources of the effectiveness data (single study, review or opinion). In the case of single studies it is expected that the authors also report details about the study design and the analysis of the data acquired. In case that effectiveness data were acquired from a review of previously published studies, review methods (search strategy, inclusion criteria etc.)

need to be reported. All these questions are supposed to be applicable to model-based economic evaluations but in our case none of the authors provided such detailed information in their publication. Therefore, this section was scored very low. Most of the authors reported the references of the effectiveness data only or reported a value of the effectiveness they assumed but this was considered insufficient according to the instructions of the checklist. For item E_5 authors need to adequately report results of the effectiveness in principal measures of effectiveness (with confidence intervals), which was stated in only few cases. Item E_6 addresses adverse effects reported in the economic evaluation analysis. None of the authors reported any side-effects included in the analysis (or commented why they did not do that). E_7 is not applicable for model based studies, as stated in the checklist. In general, the effectiveness section of the checklist lowered the average transferability score to a relatively large extent. The reason is that guidelines dictate quite detailed reporting criteria for reporting which might exceed the expectation of the level of reporting for model-based studies.

The benefit measures (B_1) were in general well reported (95% of full scores). However, in most cases authors did not report the basic method of valuation of the health states (B_2), and whose values were taken into account (B_3). Therefore, on average the scores for these two items were 22%. None of the studies specified the valuation tool used (B_4), i.e. which method was used to obtained utilities. In most of the cases only reference of the study valuing benefit measures was stated.

Costs data are known as one of the most country specific data in economic evaluations. Eleven items of the checklist assess reporting of cost data, which gives this section a great meaning. The following items were very well reported: costs included in the final calculation (C_1), sources of unit price data (C_4), unit prices for resources (C_5), price year (C_7) and currency of costs (C_9). However, only a minority of the articles reported methods used to measure components of the costs (e.g. microcosting) (C_2), sources of resource consumption data (C_3) and statistical analysis of the cost data (C_{11}). The item about the time horizon of each element of the cost analysis (C_8), and the item about separately reporting costs and quantities (C_6) were not applicable, because we had costs data reported directly per cancer patient. The item about currency conversion rate (C_{10}) is applicable

only when the data about costs are taken from the setting with another currency. This question was applicable in only five studies and was reported only in two cases.

Discounting of the costs and benefits were very well reported (D_{1-3}). Authors pointed out the comparison between discounted and undiscounted results to a lesser extend (D_4).

In item S_1 the analysis of variability from place to place was assessed. This item we understood as only applicable for multinational studies. Only in one study variability from place to place was analysed and taken into account.

The last item (O_1) of the checklist inquires if the authors of the study discussed any caveats regarding the generalizability of their results. This way a potential user can get the information (or at least opinion) about the relevance of the study at first hand. Only a quarter of scored studies contained this kind of information.

As already mentioned, the biggest disadvantage of the EURONHEED checklists appeared that that the interpretation and scoring process are very subjective. In addition, some items might be difficult to be applied in some health care interventions. Another point is that the items are not specifically weighted, which means that every item is equally important, although the authors of the method pointed out that number of items on certain topic already reflects some weighting (11 items in costs section, 7 questions about effectiveness etc.).

5.2. Comment on the scoring results

The general observation is that the total scores were higher for the subchecklist compared to the full checklist. This means that studies performed better on items most critical for transferability. Namely, the subchecklist was originally formed by selecting items that were deemed most important for transferability. None of the studies, however, scored all points on the full checklist and only one reached all points in the subchecklist. The highest score was 79 and 100 for the full checklist and the subchecklist, respectively. The variation of the total score among studies was high. Studies with lowest reporting quality had as low as 23 points and 42 points for the full checklist and the subchecklist, respectively. It can

thus be concluded that studies had in general a moderate quality of reporting items relevant for transferability.

As shown in Figure 8, reporting was weakest for effectiveness (E), benefit measures (B), and an item focusing on how the researchers addressed the issues of generalizability of results (O). Although the score was low also for the question relating to the quantitative and/or descriptive analysis conducted to explore variability from place to place (S), the question was applicable only in 3 studies. In case of effectiveness data (E), studies mostly reported references from which the data were obtained or the effectiveness that was assumed in calculations. However, the checklist addresses the reporting of effectiveness with more scrutiny and requires details of e.g. search strategy, inclusion criteria for clinical studies to be reported. Hence, the scores for this section were low. On the other hand, some other questions of the checklist are not specific enough for modelling studies (as mention above), especially the section relating to modelling (M). Namely, studies could score similarly for this section but there were clearly big differences in the details provided.

Moreover, it is difficult to interpret the total score and differences between studies. There are no instructions given by the authors of the checklist on what scores can be considered as high, moderate, or low, and what is the cut-off point to decide whether the economic evaluation is transferable or not. It is also not known whether the score is a valid summary measure. It may be that the study which scores fewer points is easier to transfer to other settings than some study with a higher score. It is also not known what the properties of the score scale are. For example, a total score of 80 does probably not indicate that the study is twice better transferable than a study with a score of 40.

We consider that the summary index represents a relatively rough measure of transferability. The scoring is subjective, especially with questions that are not specific enough, and the interpretation of the total score is undefined.

5.3. Comparison of EURONHEED checklist to other methods

None of the published methods for assessing transferability of economic evaluations has been widely tested so far. Furthermore, we made no direct comparison between the different methods for assessing transferability. However looking at the published studies we can draw some initial conclusions.

Welte's method was tested by Knies et. al.. (18) They applied the model of Welte for testing the transferability of foreign cost-effectiveness evaluations to the Netherlands. Using Welte's model they achieved better results for the cost prediction than when the foreign results would be applied straightforward in the decision country, but the effectiveness prediction was less accurate. They pointed out that only methodological characteristics could be assessed without using extra information outside the articles, and that judging health care and population characteristics seemed to be very complex. They criticized the third general knock-out criterion which states that the study should possess an acceptable quality as there is no clear definition how the quality should be assessed. Further, they exposed the overlapping of some factors and the lack of attention for the transferability of effects. Another deficiency of the method is the fact that Welte's model is focused on the idea to assess the transferability of whole studies, and in case that the study as a whole is not transferable it is not possible to assess whether a section of the study could be transferred. (12, 18)

Compared to the EURONHEED checklist Welte's method considers possible transferability factors less detailed and only provides an estimation of the transferability factors in a descriptive way. In both methods factors assessing transferability are overlapping. On the other hand, Welte's chart method provides decision maker information which study should be transferred and which not, while the EURONHEED checklist gives us only the scoring value which does not have a clear-cut meaning whether the study is transferable or not, making the interpretation of the scores unclear.

In both methods authors give an instruction, that potential user should only take into account published studies possessing an acceptable quality. However, in both cases the authors do not specify what an acceptable quality of the published study is, or how it should be assessed.

Last but not least, Welte's method is designed to assess the direction of the change of ICER in another setting, while EURONHEED checklist does not provide this kind of information and focuses only on the level of reporting in the study.

One of the most important disadvantages of EURONHEED checklist is that the items are not specifically weighted, meaning that every item is equally important. Antonanzas et. al. suggested a method where the approach for assessing the transferability of studies is theoretically a combination of Welte's method and Nixon's method (EURONHEED checklist). In the method both, the principal of general knock-out criteria and scoring system were adopted. Additionally, the issue of assigning weights to the items is addressed. The questionnaire was sent to seven different HTA agencies and in this way weights of the items were determined, based on opinion of the experts. Although they applied values of weights into calculation of the General Transferability Index, the final results were comparable to the calculation performed without weighted factors. Antonanzas et. al. reported difficulties with interpreting various factors and pointed out need to develop guidance to help understanding each factor. They did not set threshold value above which a given study would be recognized as transferable. An interesting argument for that was that a fixed threshold value would not be acceptable and useful. Namely, a specific transferability index is specific to a setting. That means, that study might be very well transferable in one setting (achieving high value of scores), but would not be universally acceptable and useful in other settings. However, this method indicates the direction of further development, because it merges approaches of already published methods. We did not decide to apply this method on our case, as we did not have an option to gain setting specific assessments of weights.

5.4. Improving transferability of economic studies

Transferability issues can be taken into account either when performing or when interpreting economic evaluations (Figure 9). At the level of performing an economic assessment, transferability is affected at the **designing, analysing**, and **reporting** stage. (13) At the designing level, the selection of the comparator therapy or perspective of the study, for example, has a major influence on applicability of the study to different settings. However, at this stage the researchers are focused on their own setting and try to design a study to approximate their local setting as much as possible. It is not likely that researchers

will e.g. include or be requested to include additional comparators in the analysis that are not relevant in their local setting just for the sake of making their study more transferable. However, transferability could be improved if different local guidelines for conducting health economic evaluations were more aligned. This could be achieved at least for some of the more general topics, such as perspective considered (societal, payer), discount rate for costs and outcomes, appropriate methods to measure utilities. Currently, there is high discrepancy between the national pharmacoeconomic guidelines. (69)

The stage where the researchers should definitely consider transferability is the reporting stage. Information from economic evaluations should be reported in a transparent and comprehensive way. We believe it would be beneficial to have guidelines on quality and explicit reporting developed, specifically tailored to maximise transferability. The development of the guidelines should be based on international collaboration of experts, because this way specifities of different health systems would be easier identified and taken into account. Moreover, separate guidelines could be developed for reporting model-based health economic studies and reporting economic evaluations conducted alongside clinical trials. A similar task has already been done in the field of quality reporting of different types of clinical trials. Standards for reporting were developed separately for randomised controlled trials (CONSORT statement), systematic reviews and meta-analyses (PRISMA statement), observational studies in epidemiology (STROBE statement), meta-analyses of observational studies in epidemiology (MOOSE statement), and nonrandomised public health interventions (TREND statement). (70-74)

Based on our research work we could see that the EURONHEED checklist, which can be used also as a guideline for maximising transferability when reporting the health economic study, is not well suited for model-based studies. It addresses the main items to consider but it should be more specific in some questions. A modelling section (M) should include questions like whether the model was graphically presented, whether all pathways and possible transitions between health states were clearly stated, whether all input parameters were provided, how parameters were obtained (literature review, experts opinion), how rates from trials were converted into probabilities, whether all assumptions were listed. Although it might be considered by some people that manuscripts would become too detailed, data could be efficiently presented in an additional electronic file which can be downloaded from the journal's web site, which is already usual practice for some journals.

Transferability of the studies could be explored and assessed in details only if models used in economic evaluation would be freely available. Consequently, a potential user could easily apply specific data to his setting.

The reporting level is very much linked to the stage of interpretation of economic evaluations, where **diagnosing** and **adaption** play the main role before the policy decision is made. In the diagnosing stage, different elements of an economic evaluation are assessed on their transferability. The diagnosing stage is very important for decision makers as they have to consider whether or not data, method used and/or results of the economic evaluation can be easily applicable to their jurisdiction. When not, the study has to be **adapted** according to specific characteristics of the jurisdiction.

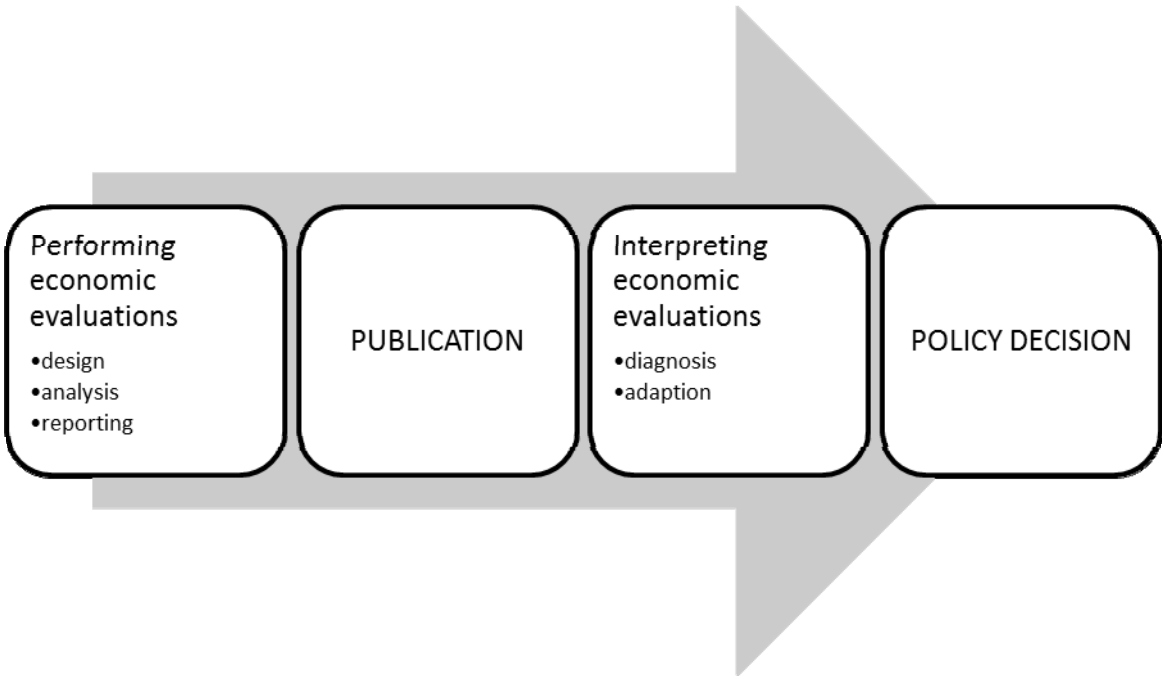

Figure 9: Transferability issues in different stages (adapted from (7, 13))

Finally, to increase transferability authors of the studies should explore the transferability of their results through performing sensitivity analyses of their results. (12) This would help to recognize the influence of certain parameters on final results. With comprehensive sensitivity analyses a potential user is provided with information how a specific parameter in the economic evaluation (which might be different between compared settings) influences the results and has to be replaced with the setting specific data. Authors should

therefore report details about the sensitivity analyses performed and provide the reader with information like what kind of analyses was performed (one-way, probablistic), which parameters were varied, which ranges were used in the analyses, potentially report the results of sensitivity analyses graphically, and most importantly, point out the most sensitive parameters and discuss them.

Welte et al. (12) recommended that for determination of the most essential adjustments of model parameters a univariate sensitivity analysis should be performed. If not all relevant study parameters can be substituted with country-specific ones, multivariate or probabilistic sensitivity analysis seems to be a promising way to quantify the uncertainty associated with a transfer. If study results cannot be transferred, the transfer of study models or designs should be investigated as this can significantly save time when conducting a new study.

5.5. Implications for decision makers

Demand for economic evaluations of health care technologies has been growing in recent decades. Studies performed at local setting are not always available therefore transferring studies performed in other setting can be the only option available. Above all, transferring data of published economic evaluation is not only time and resource saving, but sometimes the only option, when the study cannot be performed locally.

Economic evaluations cannot be transferred easily due to the differences between jurisdictions, especially between different healthcare systems. Decision makers need to be aware of the factors hindering transferability and able to assess which parameters of the evaluations are potentially relevant for their jurisdiction. This critical assessment is not needed only for costs and effectiveness data, but all critical factors that influence local value of the ICER need to be identified. Results of recent research activities in the field of transferability showed that also other elements of economic evaluations like valuing of productivity loss and utilities of health benefits differ among jurisdictions to the extents which cannot be negligible. (7) Applying values of the most important parameters without adaptation can be misleading and causing predictions of the ICER which do not hold true for the setting. Value of ICER can be either underestimated or overestimated; -both scenarios can lead to making the wrong decision.

Utilisation of the foreign published data does not mean direct adoption of the economic evaluation results, but critical insight into the evaluated health care technology and adequate adaptation of the data must be performed. Decision makers need to assess which foreign data is relevant to their setting. A recently published review of international pharmacoeconomic guidelines found that across 27 sets of guidelines, baseline risk and unit costs were uniformly considered to be of low transferability, while treatment effect was classified as highly transferable. Results are more variable for resource use and utilities, which were considered to have low transferability in 63% and 45% of cases, respectively (Figure 10). (11)

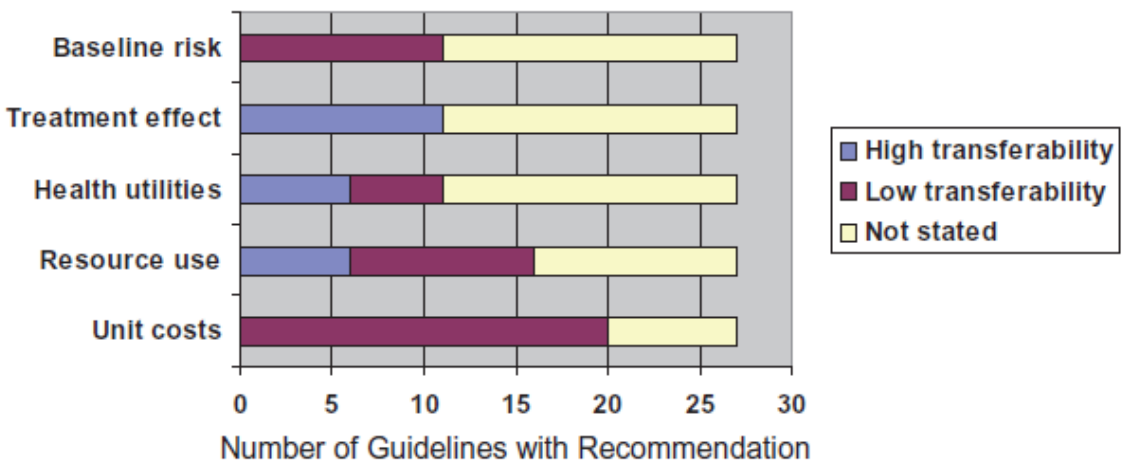

Figure 10: Transferability of different types of data according to pharmacoeconomic guidelines.

Recently published methods presented in this thesis help to assess the transferability of studies. It is important to be aware that published methods do not help to adapt the data but point out elements of economic evaluations which we should pay attention to, when applying them to our setting. Using the published methods decreases the possibility that important elements of the evaluation would be overlooked and help to recognize whether the data should be adapted to the new setting or potentially new data need to be collected.

6. CONCLUSIONS

- Summary index calculated with the help of the EURONHEED checklist represents a relatively rough measure of transferability.

- Scoring is subjective, especially with questions that are not specific enough, and the interpretation of the total score is undefined.

- Items of EURONHEED checklist should be more specified, especially for model based studies.

- Additional research of factors influencing the transferability of economic evaluations should be performed, especially their relative importance should be investigated.

- Guidelines for transparent and comprehensive reporting of economic evaluations should be developed, specifically tailored for model-based and trial-based economic evaluations.

- Future research should focus on assessing transferability of different types of data across jurisdictions (health states utilities).

7. REFERENCES

1. Ham C, Hunter DJ, Robinson R: European Observatory on health systems and policies; Evidence based policymaking. British Medical Journal (1995), 310:71–72.
2. Health Technologies Assessment international, accessed at http://www.htai.org, 15.2.2011
3. International Network of Agencies for Health Technology Assessment, accessed at http://www.inahta.org/HTA/, 15.2.2011
4. Jonsson E: Development of health technology assessment in Europe. International Journal of Technology Assessment in Health Care (2002), 18:171–183.
5. Drummond, M.F., MacGuire A: Transferability of Economic Evaluations Across Jurisdictions: ISPOR Good Research Practices task Force Report. Value in Health, 2009, 12(4): 409-418
6. Späth HM, Carre`re MO, Fervers B, Phillip T. Analysis of the elegibility of published economic evaluations to a given health care system. Health Policy (1999); 49 (3) : 161–177.
7. Knies S: Transparency of Transferability, Diagnosing international aspects of economic evaluations of health care technologies, ISBN 978 94 6159 0237 7, University of Maastricht, 2011, 10-15.
8. Boulenger S, Nixon J, Drummond M, Ulmann P, Rice S, Pouvourville G. 2005. Can economic evaluations be made more transferable? European Journal of Health Economics (2005) 4(6): 334–346.
9. Antonanzas F., Rodriguez-Ibeas R, Juarez C, Hutter F, Lorente R, Pinillos M: Transferability indices for health economic evaluations: methods and applications, Health Economics (2009)18(6): 629–643
10. Drummond M, Barbieri M, Cook JR, Glick H, Lis J, Malik F, et al., Transferability of economic evaluations across jurisdiction: ISPOR good research practices taskforce report. Value in Health (2009), 12(4): 409-418.
11. Barbieri M, Drummond M, Rutten F, Cook J, Glick HA, et. al., What Do International Pharmacoeconomic Guidelines Say about Economic Data Transferability, Value in Health, (2010), 13 (8),1028–1037.

12. Welte R, Feenstra T, Jager H, Leidl R. 2004. A decision chart for assessing and improving the transferability of economic evaluation results between countries. Pharmacoeconomics 22(13): 857–876.
13. Drummond M, Manca A, Sculpher M, Increasing the generalizability of economic evaluation: recommendations for the design, analysis and reporting of studies. International Journal of Technology Assessment in Health Care, (2005), 21(2):165-171.
14. Mason JM, Mason AR: The Generalisability of Pharmacoeconomical Studies, Issues and Challenges Ahead, Pharmacoeconomics (2006), 24(10)937-945.
15. Goeree R, Burke N, O'Reilly D, Manca A, Blackhouse G, Tarride J-E: Transferability of economic evaluations: approaches and factors to consider when using results from one geographic area for another. Current medical research and opinion (2007), 23(4):671-682.
16. Sculpher MJ, Pang FS, Manca A, et al. Generalisability in economic evaluation studies in health care: a review and case-studies. Health Technology Assessment 2004;8:1–192.
17. Drummond M, MacGuire A: Methods for the Economic Evaluation of Health Care Programmes, Third edition, Oxford University Press, (2005),323-335.
18. Knies S, Ament AJHA, Evers SMAA, Severns JL: The Transferability of Economic Evaluations: Testing the Model of Welte. Value in Health, (2009), 12(5),730–738.
19. Urdahl H, Manca A, Sculpher M: Assessing generalizability in model-based economic evaluation studies. Pharmacoeconomics, (2006), 24(12): 1181-1197.
20. Nixon J, Rice S, Drummond M, Boulenger S, Ulman P, De Pouvourville G, Guidelines for completing the EURONHEED transferability information checklist. Europena Journal of HElath Economics, (2009), 10(2):157-165.
21. Antonanzas F, Rodriguez-Ibeas R, Carmelo J, Hutter F, Lorente R, Pinillos M: Transferability indices for health economic evaluations: methods and applications. Health Economics, (2009), 18(6):629-643
22. Cox T, Cuzick J: HPV DNA testing in cervical cancer screening : from evidence to policies, Gynecol Oncol.103(1):8-11.
23. Waggoner SE: Cervical cancer. Lancet (2003); 361:2217-25.

24. Garnett GP, Anderson RM. Factors controlling the spread of HIV in heterosexual communities in developing countries: patterns of mixing between age and sexual activity classes. Philos Trans R Soc Lond B Biol Sci. (1993);342:137–59
25. Brisson M, Van de Velde N, De Wals P, Boily MC: The potential cost-effectiveness of prophylactic human papillomavirus vaccines in Canada. Vaccine. (2007) 20;25(29):5399-408
26. Adams M, Jasani B, Fiander A: Human papilloma virus (HPV) prophylactic vaccination: Challenges for public health and implications for screening. Vaccine (2007), 25(16):3007-13.
27. Harper DM, Franco EL, Wheeler C, Ferris DG, Jenkins D, Schuind A, Zahaf T, Innis B, Naud P, De Carvalho NS, Roteli-Martins CM, Teixeira J, Blatter MM, Korn AP, Quint W, Dubin G; Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. Lancet. 2004 Nov 13-19;364(9447):1757-65.
28. FUTURE II Study Group: Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. N Engl J Med. 2007 May 10;356(19):1915-27.
29. Clifford GM, Smith JS, Aguado T, Fraceschi S: Comparison of HPV type distribution in high-grade cervical lesions and cervical cancer: a meta-analysis, British Journal of Cancer (2003), 89:101-5.
30. Olsson SE, Villa LL, Costa RL, Petta CA, Andrade RP, Malm C, Iversen OE, Hoye J, Steinwall M, Riis-Johannessen G, Andersson-Ellstrom A, Elfgren K, et al. Induction of immune memory following administration of a prophylactic quadrivalent human papillomavirus (HPV) types 6/11/16/18 L1 virus-like particle (VLP) vaccine. Vaccine (2007);25:4931–9.
31. Drummond M, MacGuire A: Methods for the Economic Evaluation of Health Care Programmes, Third edition, Oxford University Press, (2005), 247-308.
32. Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, SzucsT, Vray M: Modelling in Economic Evaluation: An Unavoidable Fact Of Life, Health Economics (1997), VOL. **6:** 217–227
33. Brennan A, Akehurst R: Modelling in Health Economic Evaluation What is its place? What is its Value?, Pharmacoeconomics (2000); 17(5):445-459.

34. Goldie SJ, Goldhaber-Fiebert JD, Garnett GP. Chapter 18: Public health policy for cervical cancer prevention: the role of decision science, economic evaluation, and mathematical modeling.Vaccine. 2006 Aug 31;24 Suppl 3:S3/155-63
35. Dasbach EJ, Elbasha EH, Insinga RP. Mathematical models for predicting the epidemiologic and economic impact of vaccination against human papillomavirus infection and disease. Epidemiol Rev. 2006;28,88-100.
36. Woodman CB, Collins SI, Young LS: The natural hystory of cervical HPV infection: Unresolved issues. Nat Rev Cancer(2007), 7(1):11-22.
37. Uwe Siebert, Gaby Sroczynski, Peter Hillemanns et al. The German Cervical Cancer Screening Model: development and validation of a decision-analytic model for cervical cancer screening in Germany. European Journal of Public Health (2006), Vol. 16, No. 2, 185–192
38. Erik J. Dasbach, Elamin H. Elbasha, and Ralph P. Insinga: Mathematical Models for Predicting the Epidemiologic and Economic Impact of Vaccination against Human Papillomavirus Infection and Disease, Epidemiol Rev 2006;28:88–100
39. Annemans L, Rémy V, Oyee J, Largeron N.: Cost-effectiveness evaluation of a quadrivalent human papillomavirus vaccine in Belgium.Pharmacoeconomics. 2009;27(3):231-45
40. Bergeron C, Largeron N, McAllister R, Mathevet P, Remy V: Cost-effectiveness analysis of the introduction of a quadrivalent human papillomavirus vaccine in France, Int J Technol Assess Health Care. 2008 Winter;24(1):10-9.
41. Boot HJ, Wallenburg I, de Melker HE, Mangen MJ, Gerritsen AA, van der Maas NA, Berkhof J, Meijer CJ, Kimman TG: Assessing the introduction of universal human papillomavirus vaccination for preadolescent girls in The Netherlands, Vaccine. 2007 Aug 14;25(33):6245-56. Epub 2007 Jun 21.
42. Brisson M, Van de Velde N, De Wals P, Boily MC: The potential cost-effectiveness of prophylactic human papillomavirus vaccines in Canada, Vaccine. 2007 Jul 20;25(29):5399-408.
43. Chesson HW, Ekwueme DU, Saraiya M, Markowitz LE: Cost-effectiveness of human papillomavirus vaccination in the United States, Emerg Infect Dis. 2008 Feb;14(2):244-51.

44. Coupé VM, van Ginkel J, de Melker HE, Snijders PJ, Meijer CJ, Berkhof J: HPV16/18 vaccination to prevent cervical cancer in The Netherlands: model-based cost-effectiveness, Int J Cancer. 2009 Feb 15;124(4):970-8.
45. Dasbach EJ, Insinga RP, Yang YC, Pwu RF, Lac C, Elbasha EH: The cost-effectiveness of a quadrivalent human papillomavirus vaccine in Taiwan, Asian Pac J Cancer Prev. 2008 Jul-Sep;9(3):459-66.
46. de Kok IM, van Ballegooijen M, Habbema JD Cost-effectiveness analysis of human papillomavirus vaccination in the Netherlands, J Natl Cancer Inst. 2009 Aug 5;101(15):1083-92.
47. Elbasha EH, Dasbach EJ, Insinga RP: Model for assessing human papillomavirus vaccination strategies, Emerg Infect Dis. 2007 Jan;13(1):28-41.
48. Ginsberg GM, Edejer TT, Lauer JA, Sepulveda C: Screening, prevention and treatment of cervical cancer -- a global and regional generalized cost-effectiveness analysis, Vaccine. 2009 Oct 9;27(43):6060-79
49. Ginsberg GM, Fisher M, Ben-Shahar I, Bornstein J: Cost-utility analysis of vaccination against HPV in Israel, Vaccine. 2007 Sep 17;25(37-38):6677-91
50. Goldhaber-Fiebert JD, Stout NK, Salomon JA, Kuntz KM, Goldie SJ Cost-effectiveness of cervical cancer screening with human papillomavirus DNA testing and HPV-16,18 vaccination, J Natl Cancer Inst. 2008 Mar 5;100(5):308-20.
51. Goldie SJ, Kohli M, Grima D, Weinstein MC, Wright TC, Bosch FX, Franco E: Projected clinical benefits and cost-effectiveness of a human papillomavirus 16/18 vaccine, J Natl Cancer Inst. 2004 Apr 21;96(8)
52. Jit M, Choi YH, Edmunds WJ: Economic evaluation of human papillomavirus vaccination in the United Kingdom, BMJ. 2008 Jul 17;337:a769.
53. Kim JJ, Goldie SJ: Health and economic implications of HPV vaccination in the United States, N Engl J Med. 2008 Aug 21;359(8):821-32.
54. Kulasingam S, Connelly L, Conway E, Hocking JS, Myers E, Regan DG, Roder D, Ross J, Wain G: A cost-effectiveness analysis of adding a human papillomavirus vaccine to the Australian National Cervical Cancer Screening Program, Sex Health. 2007 Sep;4(3):165-75.
55. Kulasingam SL, Myers ER: Potential health and economic impact of adding a human papillomavirus vaccine to screening programs, JAMA. 2003 Aug 13;290(6):781-9.

56. Mennini FS, Giorgi Rossi P, Palazzo F, Largeron N: Health and economic impact associated with a quadrivalent HPV vaccine in Italy, Gynecol Oncol. 2009 Feb;112(2):370-6.
57. Oddsson K, Johannsson J, Asgeirsdottir TL, Gudnason T: Health and economic impact associated with a quadrivalent HPV vaccine in Italy, Acta Obstet Gynecol Scand. 2009;88(12):1411-6.
58. Rogoza RM, Westra TA, Ferko N, Tamminga JJ, Drummond MF, Daemen T, Wilschut JC, Postma MJ: Cost-effectiveness of prophylactic vaccination against human papillomavirus 16/18 for the prevention of cervical cancer: adaptation of an existing cohort model to the situation in the Netherlands, Vaccine. 2009 Jul 30;27(35):4776-83.
59. Sanders GD, Taira AV: Cost-effectiveness of a potential vaccine for human papillomavirus, Emerg Infect Dis. 2003 Jan;9(1):37-48.
60. Szucs TD, Largeron N, Dedes KJ, Rafia R, Bénard S: Cost-effectiveness analysis of adding a quadrivalent HPV vaccine to the cervical cancer screening programme in Switzerland, Curr Med Res Opin. 2008 May;24(5):1473-83.
61. Thiry N, De Laet C, Hulstaert F, Neyt M, Huybrechts M, Cleemput I: Cost-effectiveness of human papillomavirus vaccination in Belgium: do not forget about cervical cancer screening, Int J Technol Assess Health Care. 2009 Apr;25(2):161-70.
62. Usher C, Tilson L, Olsen J, Jepsen M, Walsh C, Barry M: Cost-effectiveness of human papillomavirus vaccine in reducing the risk of cervical cancer in Ireland due to HPV types 16 and 18 using a transmission dynamic model, Vaccine. 2008 Oct 16;26(44):5654-61.
63. Zechmeister I, Blasio BF, Garnett G, Neilson AR, Siebert U: Cost-effectiveness analysis of human papillomavirus-vaccination programs to prevent cervical cancer in Austria, Vaccine. 2009 Aug 13;27(37):5133-41.
64. Dasbach E J, Largeron N, Elbasha E H:Assessment of the cost-effectiveness of a quadrivalent HPV vaccine in Norway using a dynamic transmission model, Expert Rev Pharmacoecon Outcomes Res. 2008 Oct;8(5):491-500.
65. Kulasingam S L, Benard S, Barnabas R V, Largeron N, Myers E R: Adding a quadrivalent human papillomavirus vaccine to the UK cervical cancer screening programme: a cost-effectiveness analysis, Cost Eff Resour Alloc. 2008 Feb 15;6:4.

66. Rogoza RM, Ferko N, Bentley J, Meijer CJ, Berkhof J, Wang KL, Downs L, Smith JS, Franco EL: Optimization of primary and secondary cervical cancer prevention strategies in an era of cervical cancer vaccination: a multi-regional health economic analysis; Vaccine. 2008 Sep 15;26 Suppl 5:F46-58.
67. Suarez E, Smith JS, Bosch FX, Nieminen P, Chen CJ, Torvinen S, Demarteau N, Standaert B: Cost-effectiveness of vaccination against cervical cancer: a multi-regional analysis assessing the impact of vaccine characteristics and alternative vaccination scenarios, Vaccine. 2008 Sep 15;26 Suppl 5:F29-45.
68. Taira A V, Neukermans C P, Sanders G D: Evaluating human papillomavirus vaccination programs, Emerging Infectious Diseases November 2004, Vol. 10, No. 11
69. International Society of Pharmacoeconomics and Outcomes Research, http://www.ispor.org/peguidelines/index.asp, Accessed: 15.4.2011
70. CONSORT. Consort Statement. http://www.consort-statement.org/, Accessed: 15.4.2011
71. PRISMA. Prisma Statement. http://www.prisma-statement.org/, Accessed: 15.4.2011
72. STROBE. Strobe Statement. http://www.strobe-statement.org/, Accessed: 15.4.2011
73. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000; 283(15): 2008-12.
74. Transparent Reporting of Evaluations with Nonrandomized Designs (TREND), TREND Statement. http://www.cdc.gov/trendstatement/, Accessed: 15.4.2011

8. APPENDIX

Figure 11: Welte's decision chart